NASA Reference Publication 1039

Synthetic Aperture Radar/Landsat MSS Image Registration

JUNE 1979

NASA

NASA Reference Publication 1039

Synthetic Aperture Radar/Landsat MSS Image Registration

H. E. Maurer and J. D. Oberholtzer, *Editors*
Wallops Flight Center
Wallops Island, Virginia

P. E. Anuta, *Editor*
Purdue University
Lafayette, Indiana

National Aeronautics
and Space Administration

**Scientific and Technical
Information Branch**

1979

ACKNOWLEDGMENTS

Dr. John Rouse, formerly of NASA Headquarters, encouraged and provided support for the activities described in this report. The development of the report involved contributions from Purdue University's Laboratory for Applications of Remote Sensing (Contract NAS6-2816), IBM Federal Systems Division (Contract NAS6-2827), and the Goodyear Aerospace Corporation (Purchase Order No. 70868), as well as from NASA Wallops Flight Center and its support contractors.

The staff of Goodyear Aerospace Corporation who contributed to this publication include B. Manning and G. LaPrade. This multidisciplinary effort depended on Goodyear for tutorial help on SAR technology, as well as for help on the evaluation of SAR image formation. Goodyear contributed the sections on *Principles of SAR Operation* and *Appendix A*. Goodyear also contributed to other sections such as *Forming a Digital SAR Image* and *Sources of SAR Aircraft Imagery*.

The staff of IBM FSD who contributed to this publication include S. Murphrey and R. Bernstein. IBM also included some results of additional related studies performed under its Independent Research and Development (IRAD) program. IBM contributed the sections on *Landsat MSS Image to Map Registration*, *Landsat Processing Evaluations*, and *Sources of Landsat Imagery*, as well as most of the section *Landsat MSS Image to Map Registration*. IBM also contributed to other sections including *SAR/Landsat Data Merging Systems*, *Evaluation of Technical Results*, and *Appendix C*.

The staff of Purdue/LARS who contributed to this publication include P. Anuta, D. Freeman, W. Shelly, and C. Smith. LARS contributed the sections *Deconvolution Based Resampling Algorithms*, *Appendix B*, *Correcting Errors in SAR Imagery*, and *Control Point Location Approaches*. LARS has contributed to many other sections including *SAR/Landsat Data Merging Systems*, and *Appendix C*.

The staff of Wallops and its support contractors who contributed to this publication include H. Maurer, J. Oberholtzer, S. Amundson, W. Oderman, and P. Clemens. Wallops contributed to several sections including *Sources of SAR Aircraft Imagery*, *Evaluation of Technical Results*, *Correcting Geometric Errors in SAR Imagery*, and *Digitizing a Film Image Product*. Wallops also provided technical rewriting support where necessary, as well as much of the editorial function.

CONTENTS

FIGURES

TABLES

SYNTHETIC APERTURE RADAR/LANDSAT MSS IMAGE REGISTRATION

SUMMARY

A SAR/Landsat MSS data merging system study was conducted to determine the algorithms and procedures necessary to merge aircraft synthetic aperture radar (SAR) and Landsat multispectral scanner (MSS) imagery. From this system study, the design of a SAR/Landsat Data Merging System (SLDMS) emerged. The SLDMS design was tailored to an installation at the Laboratory for Applications of Remote Sensings (LARS) of Purdue University with the Landsat processing portion of the SLDMS designed in more detail than the SAR processing portion. Three (3) aircraft SAR images were registered to the corresponding Landsat MSS scenes. These three (3) data sets were the subject of experimental investigations. Results indicate that the registration of SAR imagery with Landsat MSS imagery is both feasible from a technical viewpoint, and useful from an information-content viewpoint. Multisensor registration is likely to be a very desirable process in the future; however, the technology is not presently available to the user.

No new technology was discovered during the course of this investigation. Rather than developing new technology, appropriate technology had to be located and often modified for use in the system. It is estimated that new technology will be needed for effective computer-based analysis of the merged SAR/Landsat data sets.

The results of this study led to the following recommendations: (1) investigations should continue to determine the differences involved in registering satellite SAR images with corrected Landsat MSS images, and techniques for modeling these geometric distortions should be investigated; (2) a SAR/Landsat Data Merging System that can be transportable with a minimum of resources should be developed and located at a user-oriented facility; (3) investigations should be continued to determine how the addition of a registered SAR channel to the Landsat MSS channels contributes to the information extraction process; and (4) investigations should continue of the effects of high order resampling on the SAR image.

Sufficient material on precision Landsat registration (to a map) technology, SAR sensing and image formation technology, and system implementation considerations is included in this report to provide an understanding of the overall scope of the SAR/Landsat registration problem. The focus of this document is merging, not analysis. Since the average reader may have less background with the SAR than with the MSS, tutorial information of SAR data sources and SAR image formation processing is included. SAR/Landsat data sets are assembled for study and methods for merging these data types are investigated. Since satellite SAR imagery was not available during the course of the study, the investigation was based on aircraft SAR imagery of flat terrain. Radiometric and geometric distortions in this sample aircraft SAR imagery are analyzed. The first data set to be investigated was selected because of the presence of urban and certain beach and field areas, while the second data set to be investigated was selected because

of the presence of agricultural targets.

This study should be of value to those NASA programs which require multisensor imagery, to the user interested in merging and/or analyzing a SAR/Landsat data set, and to the staff of any facility which is required to provide such merged data.

INTRODUCTION

The SAR/Landsat MSS Data Merging System study described in this report was initiated in response to a need for an image registration capability for precisely registering Landsat and synthetic aperture radar imagery. Interest exists in the remote sensing community in the use of imaging sensors to augment multispectral scanner data sources in the hope of improving performance over that obtainable from either sensor alone. The Landsat series of spacecraft is helping us to learn about earth from space, and the most important instrument on Landsat is the multispectral scanner. This device provides an image of the earth 100 nautical miles wide and in a strip under the spacecraft, where each pixel from the Landsat image is 79 meters by 56 meters. Aircraft and satellite SARs can contribute to this learning process, especially if imagery from these sources can be combined. An attractive feature of the radar sensor is its all-weather capability which enables it to provide imagery for missed Landsat overpasses due to cloud cover. Since there are millions of pixels in each image with quantitative data, computer processing seems a promising method of data handling and analysis.

The objectives of the R&D activities discussed are to (1) evaluate and, where necessary, develop techniques to prepare SAR/Landsat data sets for computer processing, and to provide the earth resources community with access to this technology; (2) assemble well-calibrated SAR/Landsat data sets for demonstration purposes for several applications; and (3) analyze SAR/Landsat data sets in order to identify and quantify the contribution of the SAR channel. It was planned to acquire a Seasat-1 SAR/Landsat C MSS data set to investigate techniques for merging satellite SAR/Landsat MSS imagery and to extend the specifications of the SLDMS to include satellite SAR. The Seasat-1 satellite, an ocean-monitoring, polar orbit spacecraft, launched on June 26, 1978, from the Western Test Range near Lompoc, California, was to have provided wide area L-Band radar imagery over selected land areas. However, Seasat-1 stopped transmitting data shortly after midnight EDT Tuesday, October 10, 1978, while the spacecraft was over Australia. Some Seasat-1 SAR imagery, however, is, or will be, available.

During FY'75, NASA Wallops Flight Center demonstrated that merging aircraft SAR and Landsat imagery was feasible by merging a SAR/Landsat data set of the Chincoteague, Virginia, area. (Current experiments are based on aircraft SAR imagery; however, SARs on the other platforms will soon provide radar imagery over selected areas.) Work started at NASA Wallops on the SAR/Landsat MSS data merging problem at a low, but funded, level during 1973. As a result of the first SAR/Landsat data set generated during FY'75,

NASA Wallops was encouraged to expand activities in this technical area during FY'77.

The registration of imagery was not a new problem, so it was expected that much relevant technology existed which could be modified to fit the specific case. Several computer programs existed or were planned, that would register Landsat MSS imagery to a map; e.g., one program was on the Goddard Space Flight Center MDP for registering Landsat C imagery to a map reference. How best to take advantage of this existing technology, to modify it, and then to provide user access? No suitable facility existed to which a user could bring raw Landsat MSS and SAR imagery of the same scene and, for a reasonable cost, have the imagery registered to each other and to a map at user specified accuracy and pixel spacing. Recommendations were made for a system to register SAR imagery to a Landsat reference, with implementation of the system focused on Purdue/LARS. LARS was selected as a representative site because of their expertise in the handling and processing of digital remote-sensed imagery--especially Landsat--and, also, because it is a user-oriented facility. Insufficient resources are available, however, to complete the implementation during the time frame indicated.

SYMBOLS

A	tangent of skew angle introduced to minimize registration error for an image
A/D	analog to digital
AMS	attitude measurement system
ASP	advanced signal processor
B_d	bias correction for detector d (in an MSS)
BIL	band interleaved (MSS image data tape format)
BSQ	band sequential (MSS image data tape format)
c	speed of light, m/second
CCT	computer compatible tape
CLA	control location algorithm
CM_j	mean value of intensity for pixels in column j in a digital image
CPU	central processor unit
CRT	cathode ray tube (output device)
D	number of detectors contributing to a digital image; sampling grid spacing; displacement of resampled grid line from original grid along an original grid line
DEL X	difference between the observed line number of a control point and its value calculated in a resampling/geometric correction algorithm. DELX = XOBS - XNOM
DELY	difference between the observed column number of a control point and its value calculated in a resampling/geometric correction algorithm. DELY = YOBS - YNOM
DX	difference between the pixel line number of a control point as observed on an image and its calculated position using a registration algorithm, measured in fractions of a pixel or meters.
DY	difference between the pixel column number of a control point as observed on an image and its calculated position using a registration algorithm, measured in fractions of a pixel or meters.
EDC	EROS Data Center, USGS Facility at Sioux Falls, SD.
ERIM	Environmental Research Institute of Michigan
F_c	ratio of low pass filter maximum frequency to the Nyquist frequency for a Fourier Transform of an image.
f_d	Doppler frequency shift due to platform motion.
$fi(y_j, x_j)$	elements of the matrix W, products of powers of x and y evaluated at the control points j, used in an example of a fifth order polynomial registration algorithm.
FFT	fast Fourier transform
FM	frequency modulated
FMASDE	frame mean and standard deviation equalization
g	true ground separation of two radar targets

G	radar antenna pattern gain
G_d	gain correction for detector d in an MSS
G(θ)	antenna pattern
GC	geometric correction
GCCP	geodetic check control point
GCP	geodetic control point
GDATA	image (or plot) output commands
GEMS	Goodyear Electronic Mapping System
H or h	altitude of sensor, terrain height clearance, m
H_1, H_2..	coefficients of polynomial for least squares fit in image registration algorithm
HIRAN	high precision SHORAN
HOM	Hotine oblique Mercator map projection
I(i,j)	intensity value of ith sample in the jth line
I Channel	in-phase receiving channel (SAR) used with Q (quadrature) channel to determine magnitude and phase of incoming signal
I_o	intensity value of pixel at original grid point contiguous to resampled grid line; this point is a distance D from the resampled grid line
I_1	intensity value of pixel in original grid system contiguous to resampled grid line; with d as the original grid spacing, the distance of this point from the resampled grid line is d-D
I_2	intensity value of pixel in original grid system contiguous to pixel I_1, and in line with pixels I_o, I_1, I_{-1} etc.
I_3	intensity value of pixel in original grid system contiguous to pixel I_3 and in line with pixels I_o, I_1, I_{-1}, I_2, etc.
I_{-1}	intensity value of pixel in original grid system contiguous to pixel I_o and in line with pixels I_o, I_1, etc.
I_{-2}	intensity value of pixel in original grid system contiguous to pixel I_{-1} and in line with pixels I_o, I_1, I_{-1}, etc. (in resampling algorithm)
I_{out}	intensity value of pixel after resampling
INS	inertial navigation system
K	Boltzmann's constant
I/O	input and output
Lp	length of physical antenna, m
L_s	length of synthetic array, m
LARS	Laboratory for Applications of Advanced Remote Sensing (at Purdue University)
LCC	Lambert conformal conic map projection
LLC	line length correction
LM_j	mean value of intensity for pixels in line j in a digital image
m	meter

M	mean value of intensity averaged over N pixels and D detectors
M_d	mean value of intensity averaged over N pixels from detector d
MDP	Master Data Processor
MIST	multispectral image storage tape
MSS	multispectral scanner
N	number of pixels from detector d in sample
NCOL	number of columns in a digital image scene
NF	system noise figure
NLN	number of lines in a digital image scene
NN	nearest neighbor
O	corrected intensity for a pixel recorded from detector d
P	sum of intensity of pixels 1 to N from detector d, summed for D detectors
P_d	sum of intensity of pixels 1 to N from detector d
P_t	transmitted power
PRF	pulse repetition frequency
PS	polar stereographic map projection
Q Channel	quadrature channel, phase sensitive receiving channel for SAR with reference phase set at 90¼ to the I or in-phase channel
Q_d	sum of squares of intensity of pixels 1 to N from detector d
r	turning radius of aircraft, m
R	slant range (radar to target); matrix consisting of two columns where each column contains the coefficients of a polynomial H_i and V_i respectively
R_g	ground range, m
R_{go}	ground range to target at which there is no error because of cross-track motion
R_o	slant range normal to the flight path, m
RCP	relative control point
RMS	root mean square
RSS	root-sum-square i.e. square root of the sum of the squares - commonly of the differences between observed and computed components of a vector; thus the RSS gives the magnitude of the difference.
S	distorted along track imaged length for SAR on a turning platform
s	distance travelled by SAR platform as it records the distorted length S, m
SAPPHIRE	Synthetic Aperture Precison Processor High Reliability (An Air Force SAR R and D Project).
SAR	synthetic aperture radar
SHORAN	short-range aid to navigation
SIRA	space imaging radar
SLDMS	SAR Landsat Data Merging System

SM	standard mean value for brightness of radar image pixels
SOM	space oblique Mercator map projection
SSDA	sequential similarity detection algorithm
STALO	stable local oscillator
STC	sensitivity time control
T	interpulse period of SAR; absolute temperature; and matrix made up of horizontal and vertical pixel coordinates of control points in uncorrected space; and as a superscript following a matrix indicates the transpose of that matrix.
TBD	to be determined
TM	tape mark used in some CCT formats
TR	temporal registration
u	new coordinate (with v) of image point from polynomial least squares fit algorithm
u_i	coordinate (with v_i) of i^{th} control point in uncorrected space.
UTM	universal transverse mercator (map projection)
V	velocity of SAR platform - ground speed, m/sec
v	new coordinate (with u) of image point from polynomial least square fit algorithm
v_i	coordinate (with u_i) of i^{th} control point in uncorrected space
V_{id}	intensity recorded for i^{th} pixel from detector d
V_x	cross track SAR platform velocity
V_y	platform velocity for a SAR system; intended platform velocity for a SAR system (in the presence of cross track and vertical velocity components)
V_z	vertical SAR platform velocity
V_1, V_2..	coefficients of polynomial for least squares fit in image registration algorithm
W	matrix made up of powers of x_i and y_i evaluated at each control point, where this matrix relates the uncorrected space coordinates u_i, v_i to the corrected space coordinates
W_s	linear resolution of synthetic aperture, m
WRS	world reference system
X	coordinate in original image to be registered; distance along flight line
X'	skewed coordinate with r^1 introduced to minimize registration error for an image
x_i	coordinate (with y_i) of i^{th} control point in corrected space
X_{ij}	brightness value of pixel at column j line i in radar image

X(I)	vertical coordinate (pixel line number) of control point in output space image, that is, in the registered image
XEVAL	vertical coordinate (pixel line number) of control point in input space as computed from least square fit algorithm
XNOM	computed fractional line number of pixel containing control point (GCP) using an algorithm to geometrically correct the image
XOBS	line number of pixel containing control point (GCP) in image that has been geometrically corrected
XP(I)	vertical coordinate (pixel line number) of control point in input space image, that is, in the image to be registered
Y	distance measured along the nominal line of flight; coordinate in original image to be registered
Y'	skewed coordinate (with x') introduced to minimize registration error in an image
y_i	coordinate (with x_i) of i^{th} control point in corrected space
Y_{ij}	brightness value of pixel at column j and line i after both cross-track and along track corrections
$\hat{Y}_{ij}$	brightness value of pixel at column j and line i after cross-track correction
Y(I)	horizontal coordinate (pixel column number) of control point in output space image, that is, in the registered image
Y_o	distance in the nominal line of flight direction between radar antenna and the ground target for which the error due to cross track motion is zero (this target has a ground range of R_{go})
Y_ψ	distance in the nominal line of flight direction between the horizontal normal to this line (X) and the target at ground range R_g that has the Doppler conical angle ψ
YEVAL	horizontal coordinate (pixel column number of control point in input space as computed from least squares fit algorithm)
YNOM	computed fractional column number of pixel containing control point (GCP) using an algorithm to geometrically correct the image
YOBS	column number of pixel containing control point (GCP) in image that has been geometrically corrected
YP(I)	horizontal coordinate (pixel column number) of control point in input space image, that is, in the image to be registered
β_p	antenna azimuth beamwidth
β_s	synthetic beam beamwidth
∇D	displacement error in radar image because of velocity error in navigation system
∇V	error in velocity measurement causing displacement error on SAR image
ε	errors in SAR image
θ	depression angle (for SAR) the range vector makes with the horizontal

θ_a	antenna yaw error of SAR.
θ_p	antenna pitch error of SAR
λ	wavelength
σ	standard deviation of intensity averaged over N pixels and D detectors
σ_d	standard deviation of intensity averaged over N pixels for detector d
σ_o	radar cross-section of target
τ	width of pulse
τ_c	compressed pulse width
τ_e	expanded pulse width
ϕ	phase of returned signal relative to local oscillator
ψ	angle of target from zero Doppler line; Doppler conical angle measured from V_y (the aircraft velocity)

SOURCES OF IMAGERY

Landsat

A multispectral scanner (MSS) is part of the payload of each of the first three Landsat spacecraft. In addition, inclusion of an MSS device in the Landsat-D payload is currently under consideration by NASA. The launch dates of the Landsat satellites are as follows:

Landsat-1	July, 1972
Landsat-2	January, 1975
Landsat-3	March, 1978
Landsat-D	Scheduled in 1981

The MSS instruments on the first three spacecraft image the surface of the earth in four (Landsat-1 and Landsat-2) or five (Landsat-3) spectral bands. This is done simultaneously through the same optical system. The first four bands operate in the solar-reflected spectral region, and the fifth band operates in the thermal (emissive) spectral region, as follows:

Band 1	0.5 to	0.6 micrometers
Band 2	0.6 to	0.7 micrometers
Band 3	0.7 to	0.8 micrometers
Band 4	0.8 to	1.1 micrometers
Band 5	10.4 to	12.6 micrometers (Landsat-3 only)

A complete description of the Landsat program, its available products, and how to acquire them may be found in Reference 1. A new version of this document is being prepared by NASA Goddard Space Flight Center (GSFC) and USGS EROS Data Center (EDC) and is expected to be available soon.

Sources for Obtaining MSS Digital Image Data

MSS computer compatible tapes (CCTs) may be purchased from EDC. If a user does not know which scenes he requires, he may request a computer geographic search to obtain a listing of available scenes in his area of interest. To place an order, to inquire about the availability of data, or to establish a standing order, a user may contact:

User Services Unit
EROS Data Center
Sioux Falls, South Dakota 57198
Phone 605+594-6511, extension 151

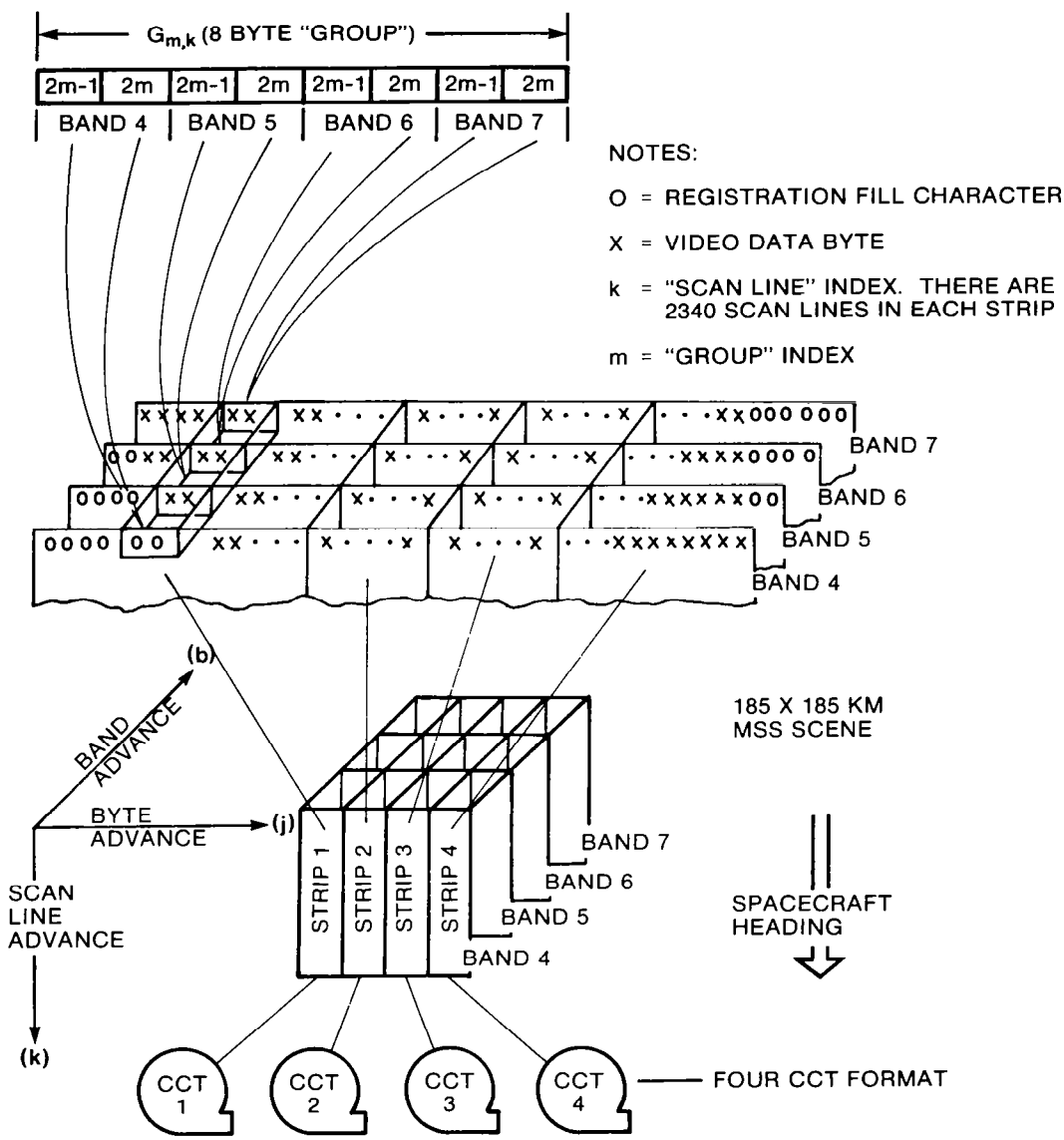


Figure 1. Interleaved format description

Data Formats for MSS Digital Data

A new processing system at EDC will become operational in 1978. Digital MSS data sets obtained prior to this are in a two-pixel-interleaved (X) format. The new system will produce data sets in either band-sequential (BSQ) or band-interleaved (BIL) format. These formats will be briefly described below.

The X format is a two-pixel, spectrally-interleaved format that has been in use since July 1972. It is used for MSS data that has been radiometrically calibrated and line-length adjusted. The data is organized into four data sets, as shown in Figure 1 . The data is contained in 8-byte "groups" that contain a corresponding pair of pixels from each spectral band. After the new EDC system becomes operational, the X format will be available for only those scenes for which EDC had previously created X format CCTs.

The BSQ format contains one data set for each spectral band. Each scan line of a band is contained in one record. It is used for either partially-processed or fully-processed data (described below).

The BIL format contains one data set per scene. Each scan line of a band is contained in one record, and these records are interleaved by band. This is shown in Figure 2 .

Scan Line 1,	Band 1
Scan Line 1,	Band 2
Scan Line 1,	Band 3
Scan Line 1,	Band 4
Scan Line 1,	Band 5
Scan Line 2,	Band 1
Scan Line 2,	Band 2
Scan Line 2,	Band 3
Scan Line 2,	Band 4
Scan Line 2,	Band 5
⋮	

Figure 2 . Band interleaved format description

Types of Processing Available

The new EDC system will provide two levels of processing for MSS data: partial processing and full processing. Partially processed data has been radiometrically corrected. That is decompression, gain, and offset adjustments have been applied to the image data. Fully processed data has, in addition, been geometrically corrected. That is, the image data has been resampled by either cubic convolution or nearest neighbor techniques to present the data in one of several possible map projections. Corrections due to spacecraft altitude ground truth, etc., have been applied.

SAR Aircraft

This section describes several sources from which synthetic aperture radar (SAR) imagery taken from an aircraft may be obtained and will indicate the formats of the data that are available to the investigator. It is written for the investigator who wants to acquire from one to several flight lines of aircraft radar data which have application to a given earth resources project. It is assumed that the investigator's interests are mainly in the output data rather than in the hardware of the radar itself. The information presented here is not written for the investigator who is interested in making various modifications to a given radar in order to obtain something other than the normally available output.

Aircraft synthetic aperture radar imagery flights are commercially available from a joint effort of Goodyear Aerospace Corporation and Aero Service Corporation and are also available from the Environmental Research Institute of Michigan (ERIM). The U.S. Air Force, Jet Propulsion Laboratory, and NASA Lyndon B. Johnson Space Center are also possible sources of synthetic aperture radar imagery flights. These sources are briefly summarized in Table 1, and the following discussion will cover each of these radar sources in more detail.

Goodyear Aerospace Corporation and Aero Service Corporation

Goodyear Aerospace Corporation and Aero Service Corporation operate a radar imaging system known as GEMS^{tm} (Goodyear Electronic Mapping System). GEMS^{tm} is an X-band high resolution radar which operates at 9.6 GHz having a wavelength of 3.12 cm. The system is installed in a Caravelle twin-jet aircraft. It provides 10 m (33 ft) azimuth resolution and 12 m (39 ft) range resolution. Horizontal polarization only is available for both the transmitter and receiver. GEMS^{tm} can be operated in either of two modes: mode A provides an image strip representing a 36 km (22 mi) wide swath from 9 km (5.5 mi) to 45 km (28 mi) on either side of the ground track, and mode B provides an image strip 36 km wide from 18 km (11 mi) to 54 km (34 mi) on either side of the ground track. The data is output in two channels each containing an 18 km swath of data for either mode A or mode B. The dynamic range of the radar is 60 db. The phase history output data is recorded on a 12.5 cm (5 in) film strip, and this data film typically is later optically correlated onto a film transparency. Goodyear can provide the investigator with the phase history data film (which due to its format will be of limited use to the typical investigator) and with an optically correlated film transparency and film print. Goodyear can also process the correlated image by what they call wide dynamic ranging which produces a color image in which high radar returns are printed in yellow and background returns are printed in red. GEMS^{tm} home airfield is located in Phoenix, Arizona. Coverage with GEMS^{tm} is available on an individual contract basis.

Table 1. CHARACTERISTICS OF RADAR SENSOR SYSTEMS

System	Center frequency (GHz)	Polarization		Resolution		Swath width (km)
		Transmit	Receive	Range (m)	Azimuth (m)	
GEMS	9.6	H	HH	12	10	36
ERIM, dual	9.45 1.315	H or V H or V	HH, HV VV, VH	9	9	5.6
AN/APQ-102A	9.6	H	HH	17	17	36
AN/APD-10	9.6	H	HH	3.7	3	18
JPL	9.6 1.215	H H or V	HH VV, HH VH, HV	20 10	20 10	20 20
JSC	9.6	H or V	VV, HH VH, HV	15	15	20

For further information about GEMS^{tm} or to make arrangements to have radar imagery flown, contact either Aero Service Corporation, Dr. E. J. Mateker, Jr., President, P.O. Box 1939, Houston, Texas 77001 (Phone: 713+784-5800) or Goodyear Aerospace Corporation, W. R. Nymann, Manager, Geosciences Services, P.O. Box 85, Litchfield Park, Arizona 85340 (Phone: 602+932-3232). Technical questions about GEMS^{tm} should be directed to Aero Service Corporation, Homer Jensen, Vice-President, 4219 Van Kirk Street, Philadelphia, Pennsylvania 19135 (Phone: 215+533-3900) or Goodyear Aerospace Corporation, L. C. Graham, Manager, Research and Development, P.O. Box 85 Litchfield Park, Arizona 85340 (Phone: 602+932-3232).

Environmental Research Institute of Michigan

The Environmental Research Institute of Michigan (ERIM) has a dual wavelength and dual polarization synthetic aperture radar system available. The system consists of two simultaneously functioning radars (an X-band radar which operates at 9.45 GHz having a wavelength of 3.17 cm and an L-band radar which operates at 1.315 GHz having a wavelength of 22.8 cm). Ground element resolution of 9 m (30 ft) in the azimuth direction and 9 m in the range direction is obtained. Vertical or horizontal polarization can be transmitted. Both of the radars have two receivers which are configured to receive the reflected signal with a polarization parallel to and crossed to the transmitter polarization, resulting in four different channels of radar data. The depression angle for both radars may be adjusted from 9° to 90°, and the maximum range is 24 km (15 mi). Data from a ground swath of 5.6 km (3.5 mi) is recorded. The polarization isolation is 23 db for the X-band radar and 19 db for the L-band radar. The dynamic range of the radar is in excess of 50 db for both radars. The received signals (a phase history) are recorded onto 70 mm data films. A data film is later processed through an optical correlator on the ground to produce an image; this can then be stored as an image on a piece of film, or it can be output on magnetic tape by using an image dissector. A modification to include recording any one of the four output channels digitally in addition to the optical film data recorded is planned. This modification might be expanded to allow all four channels of radar imagery to be recorded digitally in flight.

The ERIM radars along with some other sensors are currently being installed in a Convair 580 aircraft owned by the Canada Centre for Remote Sensing (CCRS) in Ottawa. While aboard the CV-580 the ERIM radar system, which is normally on an ERIM aircraft housed at Ypsilanti, Michigan, will be based at Ottawa; however it will also be operated out of various locations all over Canada. ERIM will operate the radars from the CV-580 for a 9 month period beginning in June 1978, during which time the Canadian government will be sponsoring many data gathering flights. The radar and other sensors will still be available to investigators in the United States on an individual contract basis between the investigator and ERIM.

For additional information and to make arrangements to have radar imagery flown, contact Robert F. Rawson, Environmental Research Institute of Michigan, P.O. Box 618, Ann Arbor, Michigan 48107 (Phone: 313-994-1200).

U.S. Air Force

The three data sets of radar imagery that are discussed elsewhere within this report and that were used to merge with Landsat data were flown by the U.S. Air Force. Until very recently, the Air Force Tactical Air Command regularly flew training mission throughout the country in F-4 aircraft that included the operation of an AN/APQ-102A synthetic aperture radar (X-band, 9.6 GHz, 3.12 cm wavelength) as part of the training exercise. It should be noted that the Air Force has decommissioned the AN/APQ-102A radars nationwide, and they are no longer available as a possible source of synthetic aperture radar imagery. Information on the AN/APQ-102A is included here since it was used to acquire the SAR imagery discussed in this report. Ground element resolutions of approximately 17 m (56 ft) are obtained with the AN/APQ-102A, and it transmits and receives signals only in a horizontal polarization. It can be operated in any one of four modes. It can record data from 0 to 18 km (0 to 11 mi) on each side of the aircraft's ground track when flown from 150 to 1500 m (500 to 5000 ft) in altitude; from 9 to 27 km (6 to 17 mi) on both sides of the aircraft at 9000 to 15,000 m (30,000 to 50,000 ft); and from 18 to 54 m (11 to 34 mi) either to the left or to the right of the aircraft at 9000 to 15,000 m. The dynamic range of the radar is 60 db. The reflected radar signals are received and recorded as a phase history on data film during flight, and the data film is later optically correlated onto a film transparency. The AN/APQ-102A data acquired by the Air Force typically was processed and evaluated. After the evaluation had been completed, the films were usually burned to recover the silver content.

Since the Air Force operated the AN/APQ-102A radars as training missions, primary concern was with the method of data collection and processing; and thus the choice of the target area to be covered was flexible. That flexibility enabled the Air Force to provide imagery to investigators who had received approval from the Department of Defense (DoD). Policy is such that requests should not place the Air Force in a competitive position with commercial interests. This restricts approval to investigators affiliated with federal government agencies and departments. In the case of AN/APQ-102A data, the Air Force provided investigators who had received DoD approval with an optically correlated film transparency and/or film print, as requested. The phase history data film was also provided if so requested by the investigator.

Although the AN/APQ-102A radars have been decommissioned, there remain other, less readily available sources of synthetic aperture radar imagery within the Air Force. There are some Air Force SAR systems based in Europe. It is possible that the Air Force would bring one of these radars back to the United States to fulfill a request for synthetic

aperture radar imagery. The requesting agency or department would be expected to defray all costs involved in the round trip of the aircraft and in the completion of the desired mission. An aircraft and the radar are on rare occasions flown to the U.S. to undergo a modification. If an investigator's request coincides with a time when the aircraft is in the States, it is possible for a request for imagery to be fulfilled without the requesting agency or department defraying the cost of transporting the aircraft.

Investigators from federal government agencies or departments interested in acquiring SAR imagery from one of the European based radars should send a formal, written request giving the desired date, location, type of imagery, and other information to INFO HQ USAFE/DOR, Ramstein AFB, Germany.

The U.S. Air Force also has some synthetic aperture radars aboard SR-71 aircraft that are based in the United States. These radars are not flown as tactical training missions and are less available than were AN/APQ-102A radar imagery flights. It is possible that the Air Force will provide agencies or departments of the federal government with SAR imagery flown with the SR-71 aircraft. Requests will be evaluated to determine whether the Air Force will be able to provide support for the specific mission being requested.

To inquire about the availability and the costs of obtaining SAR imagery from the SR-71 based radar system, a formal, written request indicating the desired date, location, type of coverage, and other information should be sent to Major John Fuller, HQ USAF/XOOIZ, Washington, D.C. 20330.

The Air Force is also currently involved in a research and development project in which they are evaluating a radar system known as SAPPHIRE (Synthetic Aperture Precision Processor HIgh REliability). The SAPPHIRE is an advanced U. S. Air Force synthetic aperture radar reconnaissance system created to demonstrate the practicality of full-scale digital processing of fine resolution radar data and to validate the extremely flexible operational concepts that are possible with digital techniques. SAPPHIRE provides an 18.5 km (11.5 mi) swath width with -30 db side lobes, more than 50 db dynamic range, and a high tolerance for aircraft motions. Horizontal polarization is used for both the transmitter and receiver. It has provisions for resolution/multiple-look tradeoff studies, for color encoded display studies, and for interfacing with a real-time change detector. The SAPPHIRE equipment is divided into airborne and ground-based subsystems. The airborne system includes a pre-processor which reduces the radar data rates only to that required for the final image, a 28-channel high-density tape recorder for recording that data, and a control console. Associated with this equipment is a modified AN/APD-10 (X-band, 9.6 GHz, 3.12 cm wavelength) sidelooking radar system and an LTN-51 inertial navigation system. Provisions are also made for a future data link which could replace the tape recorder. The ground subsystem includes both range and azimuth processors with the required wide range of tradeoff capability, a display system including both black and white and color video monitors, a digital data tape playback recorder, a digital

tape image recorder, and a laser beam recorder for hard-copy imaging. The airborne subsystem of SAPPHIRE is currently installed in a C-141 aircraft. The ground based equipment is installed at Wright-Patterson Air Force Base. In addition to recording radar data on magnetic tape, the capability exists to record it on film. Only the digitally recorded data can be processed at Wright-Patterson as part of the SAPPHIRE evaluation project. The digital radar imagery data can be stored on magnetic tape or the image can be stored photographically on film as a transparency or as a print. Phase history data recorded on film must be sent elsewhere, such as to Goodyear Aerospace Corporation or to ERIM, to be optically correlated.

The APD-10 will be made available to investigators affiliated with federal government agencies and departments as time and scheduling permit. It will be necessary for radar imagery recorded digitally to be processed on the SAPPHIRE ground based subsystem. It is the ground processor that is least available due to time and schedule requirements of the evaluation project. Investigator requests for flights which output to film do not involve the ground processor and are less likely to be rejected because of time and schedule interference, but they will still have to fit within the schedule of the aircraft. All requests will be evaluated and, if given approval, flown on a non-interference basis. Costs for obtaining the radar imagery will involve full reimbursement for all materials and operational expenses incurred for the specific mission.

Inquiries about the SAPPHIRE system should initially be made by telephone to Ralph Arenz at Wright-Patterson Air Force Base (Phone: 513+255-4113; FTS: 775-4113). Written requests for specific radar data flights should be mailed to HQ ASD/AERW, R. Arenz, Wright-Patterson AFB, Ohio 45433.

Jet Propulsion Laboratory

The Jet Propulsion Laboratory (JPL) has a modified AN/APQ-102A (X-band, 9.6 GHz, 3.12 cm wavelength) synthetic aperture radar and an L-band (1.215 GHz, 24.6 cm wavelength) synthetic aperture radar. Both radars are flown from a CV-990 NASA research aircraft that is based at NASA Ames Research Center, Moffett Field, California. Ground element resolutions of 20 m (66 ft) by 20 m can be obtained for the X-band radar, and ground element resolutions of 10 m (33 ft) by 10 m can be obtained from the L-band radar. The X-band radar transmits and receives only with horizontal polarization. The L-band radar can be configured to transmit in either horizontal or vertical polarization, and it can receive in horizontal and/or vertical polarization. Each radar records on two channels and has a maximum swath coverage of approximately 10 km (6 mi) per channel. In the case of the L-band radar, it can image a 20 km (12 mi) swath (two adjacent 10 km swaths) or it can simultaneously image the same 10 km swath recording one channel with horizontal polarization and the other channel with vertical polarization. The dynamic range of the L-band radar is approximately 40 db. The aircraft is equipped with a dual channel optical recorder for each radar which records the phase histories of the two channels onto a 12.5 cm

(5 in) film. The phase histories are later optically correlated on the ground. The optical image formed when the phase history data are correlated can be output on a piece of film to produce transparencies and prints, or it can be output onto magnetic tape by using an image dissector. In addition to the optical recorders currently in use, the capability to record the phase history radar signals on magnetic tape during flight is presently being developed. The investigator who wishes to acquire radar data from JPL should anticipate costs to cover aircraft flight time and the time of the operators.

To obtain further information concerning JPL's radar system and its availability contact Frank T. Barath, Building 180-700, Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91103 (Phone: 213+354-3550; FTS: 792-3550).

NASA Lyndon B. Johnson Space Center

NASA Lyndon B. Johnson Space Center (JSC) has a modified AN/APQ-102A (X-band, 9.6 GHz, 3.12 cm wavelength) radar which is flown in an RB-57 aircraft based at JSC. The imagery provides ground element resolutions of approximately 15 m (49 ft). The radar can transmit with either horizontal or vertical polarization, and it receives both horizontal and vertical polarization of the same swath simultaneously on two recording channels. The radar can be operated in either of two modes. One mode views from 10° to 30° to the right of the aircraft and will provide a 7.5 km (4.5 mi) image swath from about 3.1 km (2.0 mi) to 10.6 km (6.5 mi) from nadir when flying at 18,300 m (60,000 ft). The other mode views from 45° to 65° to the right of the aircraft and will provide a 20 km (12.5 mi) image swath from about 18.5 km (11.5 mi) to 38.5 km (24 mi) from nadir when flying at an altitude of 18,300 m (60,000 ft). The dynamic range of the radar is approximately 60 db. As the radar system is currently configured, the received radar signals are recorded for each channel as a phase history on 12.5 cm (5 in) data film during flight. Under a contract with Goodyear Aerospace Corporation the phase history data films are optically correlated to produce transparency and print photographic images. The option to record the received radar signal on magnetic tape during flight is being considered and may be implemented. In this case, the magnetic tape containing the phase history data would be sent to the University of Texas to be processed to give a computer compatible tape containing the image data.

Investigators who have questions concerning JSC's radar may contact Ted K. Sampsel, National Aeronautics and Space Administration, Experiment Systems Division, Sensor Systems Branch, Lyndon B. Johnson Space Center, Houston, Texas 77058 (Phone: 713+483-2846; FTS: 525-2846). To make arrangements to have radar imagery flown, an investigator should contact Olav Smistad, National Aeronautics and Space Administration, Earth Resources Program Office, Operations Planning and Requirements Office, Lyndon B. Johnson Space Center, Houston, Texas 77058 (Phone: 713+483-4691, FTS: 525-4691).

An investigator who wants to acquire synthetic aperture radar imagery from any of the sources described above should contact the desired source as far in advance as possible.

This will allow arrangements and flight scheduling to be completed so the SAR flight can be made on the desired date.

Data Bank of Existing Synthetic Aperture Radar Imagery

If a user wants radar imagery of a certain area and does not require that it is flown on a specific date and at a specific time of day, imagery which has already been acquired might be sufficient. There are not many sources for historical radar imagery, however, Goodyear Aerospace Corporation operates a data storage and retrieval center for the U.S. Air Force. Radar imagery from certain Air Force programs is stored in this data bank. There are about 15.5 million km^2 (6 million mi^2) of radar imagery of various portions of the United States. X-band radar imagery at a scale of 1:100,000 having no better than 3 m (10 ft) resolution has been declassified by the military, and the majority of the imagery may be released to non-military organizations. The radar imagery may be obtained at resolutions of 3, 6.1, 12.2, or 15.2 m (10, 20, 40, or 50 ft) and at scales from 1:100,000 to 1:600,000. The radar imagery is catalogued under several categories including geographic coordinates. A map of the United States showing approximate coverage for all available radar imagery will be supplied upon request. When imagery is requested of a certain area, the geographical coordinates and a radius are entered into a computer that results in a printout listing all available radar imagery covering the given coordinates. The printout and instructions on how to use it are then sent to the requester, who must plot the coordinates from the printout and determine which of the passes are desired. The requester then sends a letter to Goodyear Aerospace requesting a specific pass or passes. A statement explaining what the imagery will be used for should be included with the request. A contact transparency and a paper print are reproduced and sent to the requester. The cost for reproducing the radar imagery is $82 per pass. Each reproduction is approximately 12.5 cm x 38 cm (5 in x 15 in) and covers between 259 and 6475 km^2 (100 and 2500 mi^2) depending on the resolution and scale of the imagery. Additional prints and enlargements are available, and prices will be provided when requested.

Inquiries and requests should be mailed to Mr. C. A. Anderson, Building 13-2S, Goodyear Aerospace Corporation, P.O. Box 85, Litchfield Park, Arizona 85340.

Goodyear Aerospace Corporation does not provide radar imagery in a digital format. However, the film image product can be digitized elsewhere on any number of film digitizers; this applies to radar imagery obtained from the data bank and to imagery obtained from any source which is provided as a film product. Further information about digitizing a film image product is given in this report in the section entitled "Forming a Digital SAR Image".

Real Aperture Radar Systems

In addition to synthetic aperture systems which use the forward motion of the aircraft to create the effect of an aperture which is much longer than the actual antenna (hence the name synthetic aperture), there are also real aperture radar imaging systems

which are available to an investigator. For real aperture systems the ratio of the wavelength to the actual antenna size determines the angular spatial resolution. Such systems are generally useful only at short wavelengths and relatively short range; however, the received signal of a real aperture system can be displayed without further processing, whereas the synthetic aperture radar systems require additional processing which can be time consuming and can involve large amounts of computer memory on digital processors. There will be no detailed discussion in this report of real aperture radar systems, but several of these systems will be briefly mentioned.

Motorola Aerial Remote Sensing, Inc.

Motorola Aerial Remote Sensings, Inc. (MARS) has an APS-94 real aperture radar (X-band, 9.24 GHz, 3.24 cm wavelength) imaging system that is available on an individual contract basis. MARS also maintains a file of some radar imagery which they have acquired. An investigator can obtain answers to any questions concerning the radar system and the services provided by MARS by contacting Ron Gelnett or Arney Stensrut, Motorola Aerial Remote Sensing, Inc., 4350 East Camelback Road, Phoenix, Arizona 85018 (Phone: 602+994-6450).

University of Kansas Remote Sensing Laboratory

The University of Kansas Remote Sensing Laboratory has a real aperture radar that is used in various research projects. It is an X-band (10.0 GHz, 3.0 cm) radar that transmits and receives only in vertical polarization. The received radar signal is usually recorded on video tape during flight and displayed on a television monitor when back on the ground, however several options are possible. The staff of the Remote Sensing Laboratory is interested in research projects which involve their graduate program, and they are not interested in providing radar imagery solely as a service that can be purchased. Inquiries should be directed to Dr. Richard Moore, University of Kansas, Remote Sensing Laboratory, 2291 Irving Hill Road, Lawrence, Kansas 66045 (Phone: 913+864-4836).

NASA Lewis Research Center

NASA Lewis Research Center (LeRC) has two APS-94 real aperture radars (X-band, 9.245 GHz, 3.24 cm wavelength) which are on loan from the U.S. Army. LeRC developed a system that uses the radar for ice mapping in the Great Lakes with rapid processing of the data so that timely information about ice flows can be provided to shipping. The system was developed for the U.S. Coast Guard. It is currently installed in a Coast Guard aircraft and is being used on an operational basis. The system hardware will eventually be transferred to the Coast Guard after which it will be their responsibility. LeRC has more recently used the APS-94 in an experiment studying ice properties in the Arctic. Inquiries about the LeRC radar system should be directed to Richard T. Gedney, National Aeronautics and Space Administration, Lewis Research Center, 21000 Brookpark Road, Cleveland, Ohio 44135 (Phone: 216+433-4000, ext. 209; FTS: 294-6209).

Available SAR/Landsat Registered Data Sets

There are three data sets containing registered SAR/Landsat imagery that are available on computer compatible tapes to a user from the Laboratory for Applications of Remote Sensing (LARS), Purdue University. These three data sets are discussed throughout various other sections in this report. They are all areas located on the Delmarva Peninsula. Data set 1 is centered around Chincoteague, Virginia; data set 2 is centered around Salisbury, Maryland; and data set 3 is centered around Cambridge, Maryland. A user may obtain copy tapes or may use the LARS multispectral image processing system (LARSYS) to work with the registered data sets. Additional information is available by contacting the Laboratory for Applications of Remote Sensing, Purdue Research Park, 1220 Potter Drive, West Lafayette, Indiana 47906 (Phone: 317+749-2052). Table 2 gives the information needed to identify a given registered data set on the LARS system.

SAR Satellite

This section describes sources from which synthetic aperture radar (SAR) imagery taken from an orbiting satellite may be obtained and will indicate the formats of the data that are planned to be available to the investigator. It is assumed that the investigator's interests are mainly in the output data rather than in the hardware of the radar itself.

At the time of this writing, there are two SAR's scheduled to operate in orbit- - SEASAT-A scheduled for launch in June 1978 and Space Imaging Radar A (SIR A) scheduled for launch in July 1979. The SEASAT SAR will only operate when in-line-of-sight with one of five data link stations. Repeat coverage can be obtained on 152-day intervals. SIR A data will be recorded on an on-board optical recorder. Data collection will be limited by recorder film capacity which is about six hours.

The SEASAT A synthetic aperture imaging radar system is the first radar system of its kind designed for the study of ocean wave patterns from orbit. The basic requirement of this system is to generate continuous radar imagery with a 100 km ground (ocean) range swath with 25-meter ground-range resolution from an orbital altitude of 800 km. The L-band (1.2782 GHz, 23.5 cm wavelength) system transmits and receives only with horizontal polarization.

The antenna look angle is centered at 20 degrees off nadir. At 800-km altitude, this gives a ground-range swath between 240 km and 340 km. The received signals via the data link are digitally encoded to five bits per word, buffered and recorded on a high-density digital tape recorder. The data recorder has a capability to record a maximum of 15 minutes of data, which corresponds to one station pass. The high-density digital tape is then used at the ground central data processor to convert the radar video signals in digital form to a radar image.

In the first year of operation, it is expected that 15,000 to 20,000 minutes of data will be recorded. At the time of this writing, there are plans to process only about 2,600

Table 2. MERGED SAR/LANDSAT DATA SET DESCRIPTION. DATA STORED IN LARSYS 3.1 FORMAT IN TAPE LIBRARY AT LARS

Data Set No.	Site Identifier	Date of SAR Flight	Landsat Frame/Date	LARS Data Set No.	Number of Lines	No. of Samples/Line	Pixel Size	No. of Channels	Tape No.	File No.
1	Wallops Island Virginia	August 30, 1973	1403-15132 Aug 30, 1973	73120104	3352	418	57x79M	6	3352	2
2	Salisbury, Maryland	August 22, 1976	2579-14535 Aug 23, 1976	76016404	2700	1906	25.4 x 25.4 M	5	3620	1
3	Cambridge Maryland	August 22, 1976	2579-14535 Aug 23, 1976	76016413	681	598	25.4 x 25.4 M	7	3692	1

minutes of this data onto four film strips, each containing a 25-km swath with some overlap. The image film will be 70 mm in width. The film data will include image, time tick/range marks, time code, space for photometric calibration pattern, analog system calibration wedges, and engineering data. See Figure 3 for image film format. Approximately 10 percent of the data imaged on film will also be recorded on computer-compatible tapes (CCT's).

The CCT will consist of quantized four-look imagery and header data. Format on the tape for this data is shown in Figure 4 . A list of the header label is given in Table 3 .

Table 3. CCT HEADER LABEL

1. Time code in days, hours, minutes, seconds for beginning image line
2. Orbit number
3. Latitude of first element of beginning image line
4. Longitude of first element of beginning image line
5. Latitude of first element of ending image line
6. Longitude of first element of ending image line
7. Array size
8. Pixel spacing
9. TBD

The image record units or frames contained within the CCT's will be composed of 1024 pixels by x lines, where $1024 \leq x \leq 10{,}240$.

SIR A, scheduled for launch in July 1979 like SEASAT A, is an L-band (1278.2-MHz, 0.235-meter wavelength) system that transmits and receives only with horizontal polarization. The intended swath width is 50 km, with range and azimuth resolution of 40 meters. On-board data recording will permit up to six hours of data collection anywhere around the world. The 50-km swath is centered at 47 degrees off nadir. Nominal altitude of SIR A is 200 km.

Algorithms for data reduction, formatting, evaluation, and analysis are to be developed prior to SIR A launch date. All the SIR A data is to be processed within six months. The original raw data will be retained at the Jet Propulsion Laboratory.

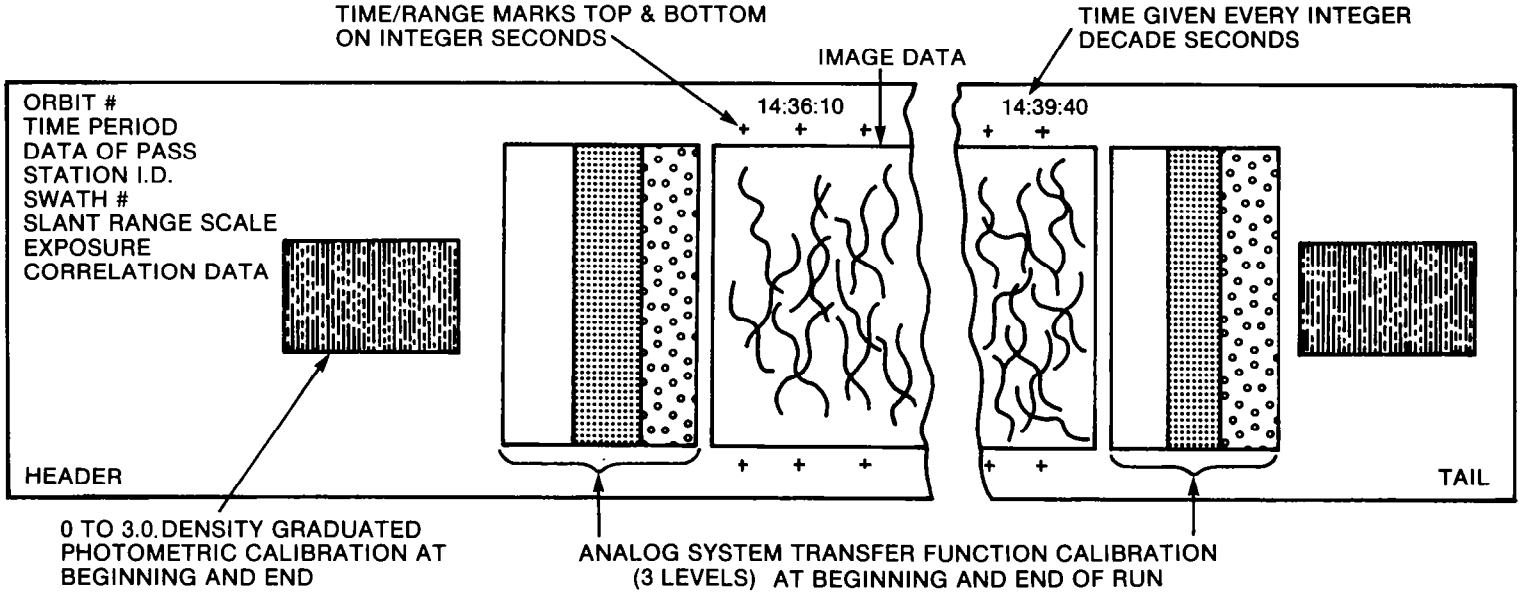


Figure 3. Image film format

THE TAPE FORMAT WILL BE DESCRIBED IN TWO PARTS. ALL HEADER VALUES WILL BE INTEGER*4 FORMAT, WHILE ALL DIGITIZED VALUES WILL BE BYTE FORMAT. 'TM' IS A TAPE MARK TO DESIGNATE THE END OF A FILE.

TAPE FORMAT

HEADER
DIGITIZED LINE 1
⋮
DIGITIZED LINE N
TM
HEADER
DIGITIZED LINE 1
⋮
DIGITIZED LINE M
TM
⋮
HEADER
DIGITIZED LINE 1
⋮
DIGITIZED LINE Z
TM
TM

N, M, Z ARE NOT NECESSARILY THE SAME.

HEADER RECORD FORMAT

DESCRIPTION	BYTES
TIME IN DAY, HOUR, MINUTES, SECONDS	16
ORBIT NUMBER	4
LATITUDE OF BEGINNING SAMPLE OF BEGINNING LINE	12
LONGITUDE OF BEGINNING SAMPLE OF BEGINNING LINE	12
LATITUDE OF BEGINNING SAMPLE OF ENDING LINE	12
LONGITUDE OF BEGINNING SAMPLE OF ENDING LINE	12
NUMBER OF DIGITIZED LINES	4
NUMBER OF SAMPLES/LINE	4
(TBD)	(TBD)

Figure 4. Computer compatible tape header and image formats

LANDSAT MSS IMAGE TO MAP REGISTRATION

Selection of Map Coordinate System

Map Projection

Map projection is a transformation of locations on the curved surface of the earth to locations on the plane surface of a map. It can be shown that it is impossible to perform such a transformation without introducing some distortion. The distortion introduced by a map projection may be in area, length, angle, or shape. Ideally, areas on a map should maintain correct proportion to areas on the earth, distances on a map should maintain true scale, directions and angles on a map should remain true, and shapes on a map and the earth should be the same. Common map projections in use today generally maintain one or two of the above ideal features while holding the distortions of the other features to tolerably low levels.

Alternate Map Projections

There are many map projections that have been mathematically defined and that are in use today. Only conformal projections, i.e., those projections which correctly preserve the shapes of small geographic features, will be considered. The Lambert conformal, the Mercator, and the stereographic projections are common conformal projections. The Lambert conformal projection is a projection from a spheroidal model of the earth to a tangent or secant cone. Mercator projections are from the earth spheroid to a tangent or secant cylinder. Stereographic projections are from the earth spheroid to a tangent plane.

The particular versions of these projections that clearly deserve consideration are

1. Lambert Conformal Conic (LCC)
2. Universal Transverse Mercator (UTM)
3. Space Oblique Mercator (SOM)/Hotine Oblique Mercator (HOM)
4. Polar Stereographic (PS)

The distortions of each of these projections are acceptable for a Landsat MSS image.

UTM Projection is Selected

The UTM map projection will be used in the SAR/Landsat Registration System (SLDMS). Although the map projections listed above are essentially equally suited for the system from a distortion point of view, the UTM projection is clearly the best choice.

Both the LCC and the UTM are widely used. Almost all of the state plane coordinate systems are either LCC or UTM. However, all federal charts are annotated with the UTM grid, as are most state charts. The UTM is officially adopted for geodetic mapping in the Soviet Union, Germany, Great Britain, and elsewhere. It is the most widely used map projection. The SOM and HOM are new map projections for which there is no existing

map base.

The UTM projection is not suitable in polar regions. The PS projection would clearly be the best projection in polar latitudes. Since scenes that are expected to be processed on SLRS are non-polar scenes, the polar stereographic projection will not be considered.

A technique not considered above would be to use raw Landsat MSS data instead of correcting it to a standard map projection. There are several reasons for not doing this. While resampling the Landsat image data, high-frequency and global geometric errors are corrected. It is also possible to obtain temporal registration between different Landsat scenes. Given that the resampling will be performed for these reasons, it is no more costly to resample to a standard map projection. Then the advantage of having the image data registered to a map is also obtained.

Radiometric Considerations

Destriping Problem Analysis

If all detectors in a given spectral band do not exhibit the same response characteristics, an undesirable phenomenon known as striping will appear in the digital image. In photographic products produced from digital data, striping can be a severe cosmetic flaw. It can, in some cases, be a hindrance to photo interpretation. Striping can be quite detrimental to multispectral classification applications because spectral signatures of the digital data are artificially altered by the striping effect. This results in reduced classification accuracy.

The ground-based radiometric calibration process used on the Landsat 1, 2, and 3 MSS data has not removed all of the detector-induced striping. Residual striping has been particularly evident in the first (0.5 to 0.6 micron wavelength) MSS band. This can be seen in Figures 5 and 6 .

Destriping Algorithms Considered

Several data-dependent techniques for destriping were investigated (Table 4). Each of these used one or more statistical parameters that are derived from the image data. The region of the image data that is used to calculate the statistical parameters is called a window. The windows considered during the investigation, as shown in Figure 7 , were rectangular areas of several sizes and shapes. All algorithms tested attempt to equalize one statistical function of the data within some window.

Figure 5. Subimage of Delaware Scene 2579-14535, Band 4 showing striping.

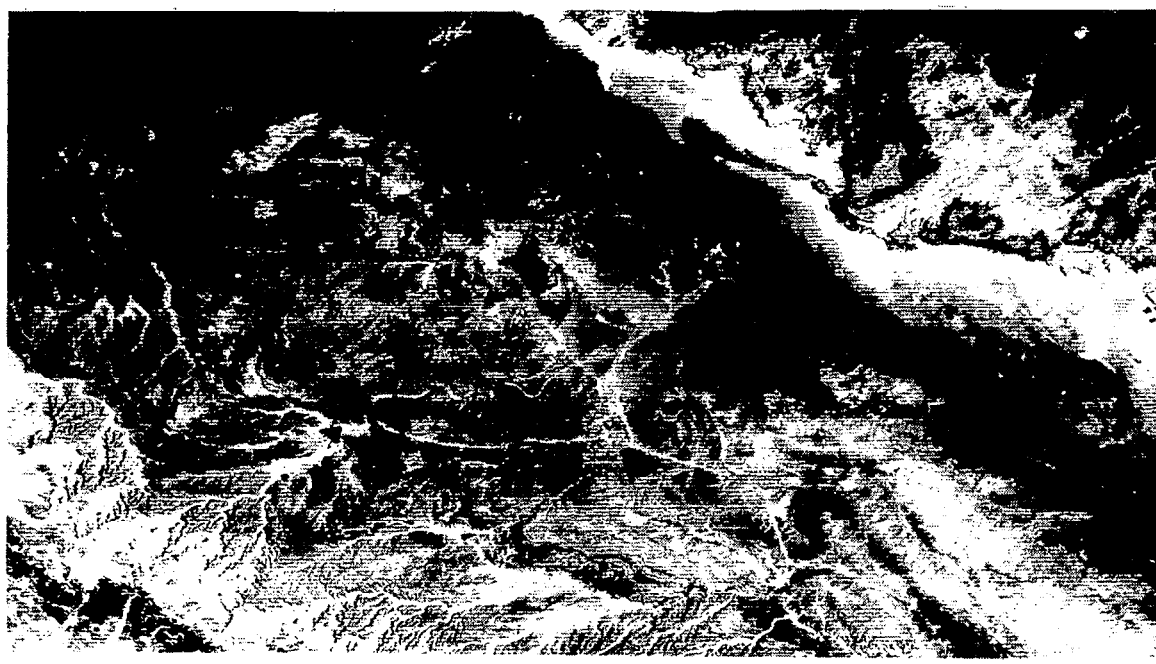


Figure 6. Subimage of Saudi Arabia Scene 1009-06523, Band 4 showing striping.

Table 4. DESTRIPING ALGORITHMS CONSIDERED

Frame Mean and Standard Deviation Equalization
Frame Mean and Standard Deviation Equalization to 1 Detector
Sweep Mean and Standard Deviation Equalization
Sweep Histogram Equalization
Local Mean and Standard Deviation Equalization
Local Averaging

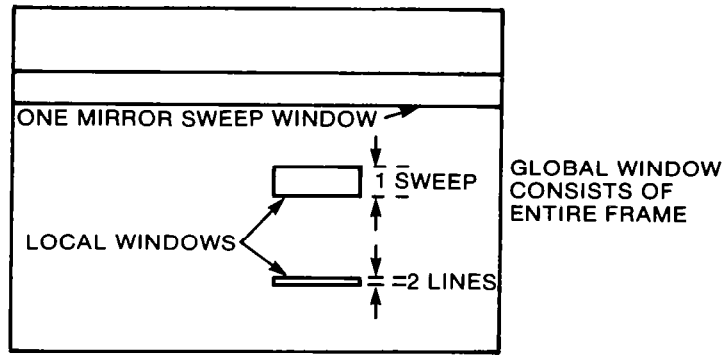


Figure 7. Window sizes used in destriping experiments

The first algorithm that was studied consists of equalizing the mean (M) and standard deviation (σ) of each detector's data over a window that is defined to be one entire frame. This equalization can be accomplished by a linear function whose coefficients are shown in Table 5. For each detector, only the sum of the pixel values and the sum of the squared pixel values are required. The result is a global function (for each detector) whose coefficients are constant during one frame of data. After destriping, the data from each detector has mean and standard deviation equal to that of the entire frame before destriping.

Table 5. MEAN AND STANDARD DEVIATION EQUALIZATION EQUATIONS

V_{i_d} = value of i th pixel from detector d

$$P_d = \sum_{i=1}^{N} V_{i,d} \qquad P = \sum_{d=1}^{D} P_d$$

$$Q_d = \sum_{i=1}^{N} V_{i,d}^2 \qquad Q = \sum_{d=1}^{D} Q_d$$

$$M_d = \frac{P_d}{N} = \text{mean} \qquad M = \frac{P}{DN}$$

$$\sigma_d = \sqrt{\frac{Q_d}{N} - M_d^2} = \text{standard deviation} \qquad \sigma = \sqrt{\frac{Q}{DN} - M^2}$$

$$G_d = \frac{\sigma}{\sigma_d} = \text{gain}$$

$$B_d = M - G_d M_d = \text{bias}$$

$$O = G_d V + B_d = \text{corrected pixel value}$$

The second algorithm is quite similar to the previous technique. It also uses frame mean and standard deviation equalization. Only one reference detector's data are corrected using the calibration wedge data before destriping. The remaining detector's data are then corrected so that they have the same mean and standard deviation as that of the reference detector. This is the approach currently used by the Canada Centre for Remote

Sensing.

The third algorithm also uses equalization of each detector's mean and standard deviation. It differs from the first algorithm in that the window over which it operates is much smaller. The window consists of all data from one sweep of the scan mirror. This technique results in a new set of linear functions (one for each detector as in Table 5 for each mirror sweep).

The fourth algorithm is also a mean and standard deviation equalization technique. However, the window involved is much more local in this method. It is a small rectangular segment of one mirror sweep. As the window is "moved" across the mirror sweep, the mean and standard deviation for each detector's pixels within the window is calculated. These are called local means and standard deviations, and they are assigned to the center of each line segment in each window. For each pixel (except those near the left or right edges of the image), the equations in Table 5 give a correction function.

The fifth algorithm uses the three steps shown in Figure 8 . The image intensity data for the entire frame is first expanded linearly to fill the interval (0,255). Then frame mean and standard deviation equalization is performed. Finally, a local averaging process is used to attempt to remove residual detector biases. This technique uses a small, two line window, as shown in Figure 7 . The image is processed in one mirror sweep segments. The first line, from detector one, is left unchanged. Each of the remaining lines in that sweep is processed to be similar to the corrected line above it. A local average of the pixels with a small window around each pixel is computed. This computation is done for the line being processed and for the already processed line immediately above it. The difference between the averages of the reference line neighborhood and the neighborhood of the line being processed is added to the intensity value of the current pixel to get the corrected intensity. This is the approach presently used by the EROS Data Center.

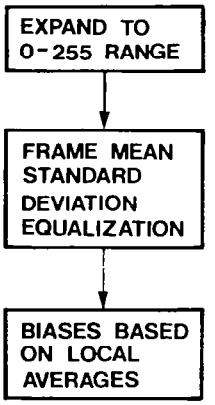


Figure 8. Local averaging methods

The sixth algorithm uses a one-mirror sweep window as a basis for histogram equalization. One detector is chosen as a reference detector and is not changed by this destriping process. Each line from the other detectors is altered to have a histogram that is similar to the line of the reference detector from the same mirror sweep. This is illustrated in Figure 9 , including the equation of the transformation.

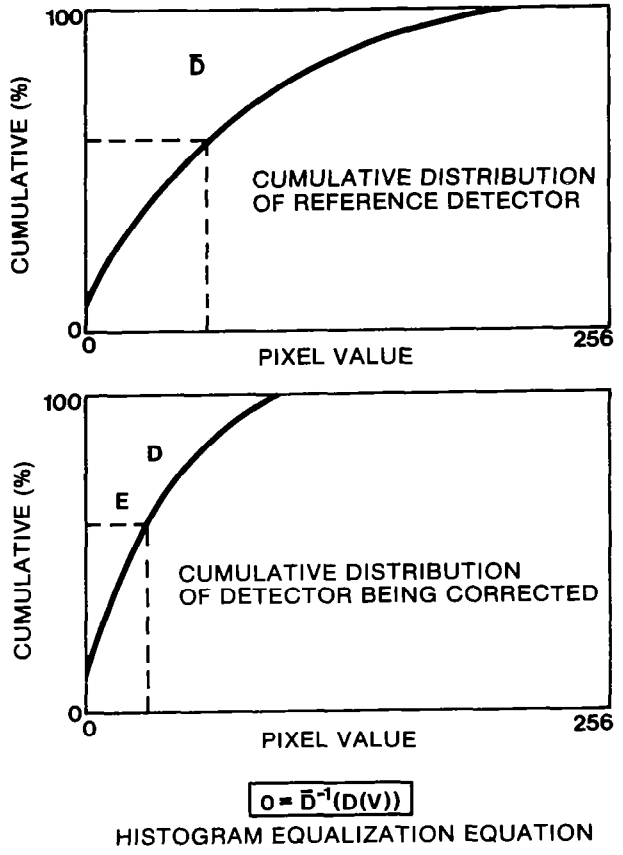


Figure 9. How histogram equalization works

Results of Destriping Algorithm Tests

A series of destriping tests were conducted using Landsat-2 MSS data as input. For this purpose, two scenes of the Washington, D.C., area (2886-14464 and 2904-14455) were obtained from NASA. These scenes contained raw, 6-bit compressed data that had not been radiometrically calibrated or altered by NASA. The CCT's for these two scenes also contained the uncalibrated, six-word calibration wedge from each pair of mirror sweeps. Uncalibrated input data was chosen for the destriping experiments in order to have maximum control over the processing that was performed on the data. An image of the uncalibrated data is shown in Figure 10 .

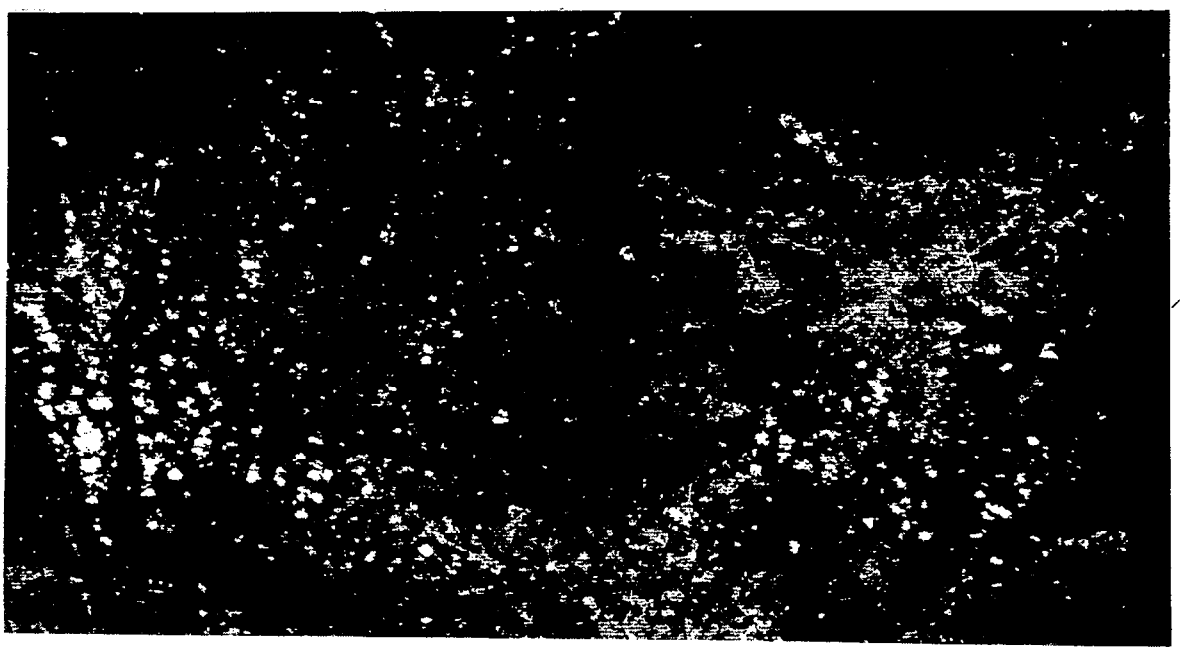


Figure 10. Subimage of uncalibrated data from Baltimore area of Scene 2904-14455, Band 4

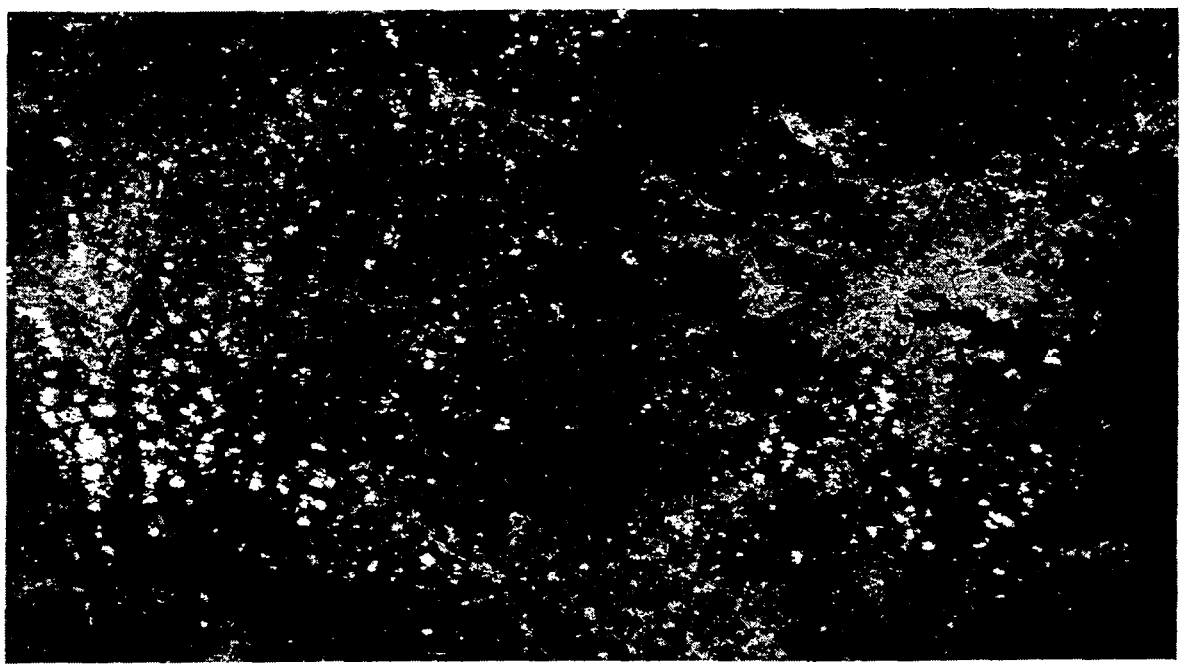


Figure 11. Subimage of NASA-Style calibrated data from Baltimore area of Scene 2904-14455, Band 4

This input data was first radiometrically corrected using the NASA decompression tables and the NASA linear regression coefficients. Gain and bias values were computed for each pair of mirror sweeps from the calibration wedges and the linear regression coefficients. The averages of all these gain and bias values were used as a global gain and bias radiometric correction function. The data corrected with this global function was used as the input data for the radiometric correction experiments. An image of this test data is shown in Figure 11 .

The results of the tests are summarized in Figure 12 . Photographic recordings of the results of applying the six candidate destriping algorithms to the test data set are shown in Figure 13 .

Algorithm	Picture Quality Comments	Implementation Cost
Raw data	Severely striped	--
NASA style Calibration	Considerably improved, moderately striped	--
Frame mean and standard deviation equalization	Further improvement, slightly striped	Low
Sweep mean and standard deviation equalization	Striping worsened, clearly unacceptable	Medium
Local mean and standard deviation equalization	Obvious local distortions, clearly unacceptable	High
Sweep histogram equalization	Striping worsened, clearly unacceptable	High
Local averaging	Further improvement, slightly striped, some local distortions evident	High

Figure 12 . Summary of destriping test results

Although frame mean and standard deviation equalization (FMASDE) shows some slight residue of striping (see Figure 13a),it is clearly the best of the six "destriped images." The algorithms that use windows of a sweep or less either increase the sweep-to-sweep striping (see Figure 13c and 13e) or introduce local distortions (see Figures 13d and 13f). Thus, FMASDE is the best choice from a data quality standpoint.

Frame Mean and Standard Deviation Equalization is Selected

FMASDE was selected for implementation in the SLRS because it produces the highest quality results of the algorithms tested and has the lowest implementation cost. It is disturbing, however, the FMASDE did not totally eliminate visible striping from the test data set (although it performed better than any of the other approaches). To investigate this further, FMASDE was applied to the raw, uncalibrated data, with the results shown in Figure 14. Virtually all the striping has been removed. This shows clearly that the difficulty in removing striping from the calibrated data is linked to the calibration process and reinforces the expectation that objectionable detector-to-detector differences would not be present in properly calibrated data.

Geometric Considerations

General Concepts

Geometric Correction (GC) and Temporal Registration (TR) are two slightly different versions of the same concept. When a digital image of an area on the earth's surface is generated by a LANDSAT MSS instrument, it will contain geometric distortions that are caused by several sources. It is desirable to modify the digital data in such a way that the geometric distortions will be removed. Geometric correction and temporal registration are two versions of a process by which the geometric distortions in an image are characterized and a mathematical transformation between the correct and distorted versions of the image is defined. The actual creation of the corrected image data set is called resampling.

The difference between geometric correction and temporal registration is the definition of what constitutes a correct image. For geometric correction, a correct image is a map. That is, a geometrically corrected image must register with a map. For temporal registration, a correct image is a previous, geometrically corrected reference image. That is, a temporally registered image must register with some other digital image. The difference is illustrated in Figure 15 . Since the reference image was geometrically corrected to register with a map, a temporally registered image nominally registers with the same map. A subtle, but important difference is that the temporally registered image has the same geometric errors as does its reference image. Therefore, temporally registered images are well suited for multispectral classification applications.

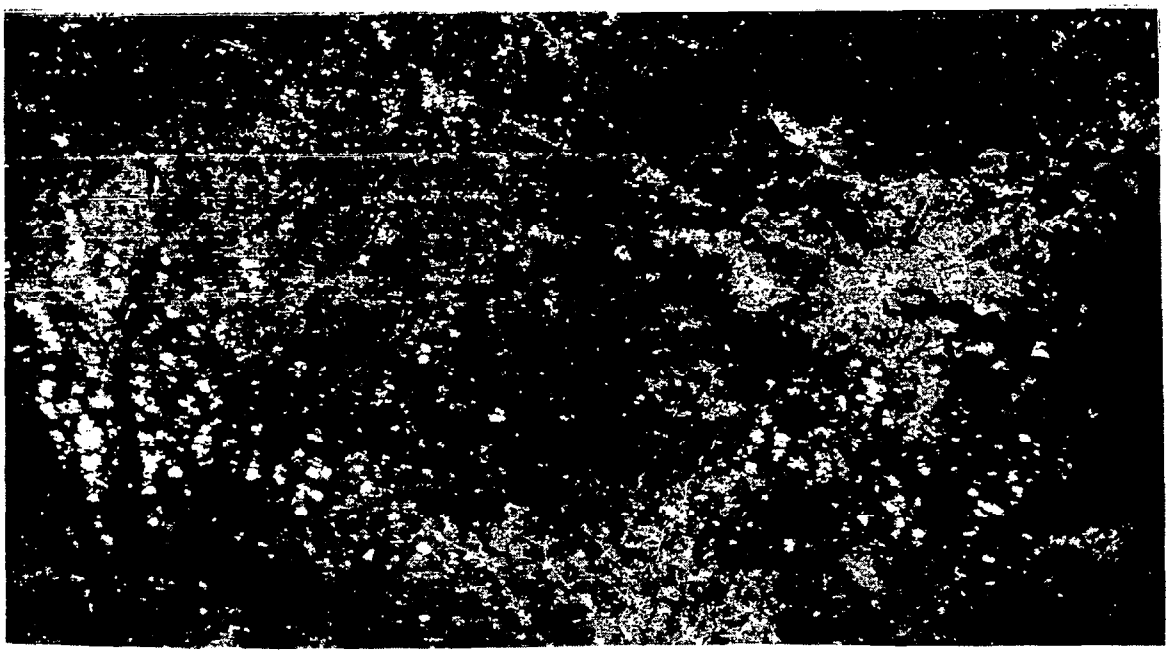


Figure 13A. Frame mean/standard-deviation equalization on Baltimore area of Scene 2904-14455, Band 4

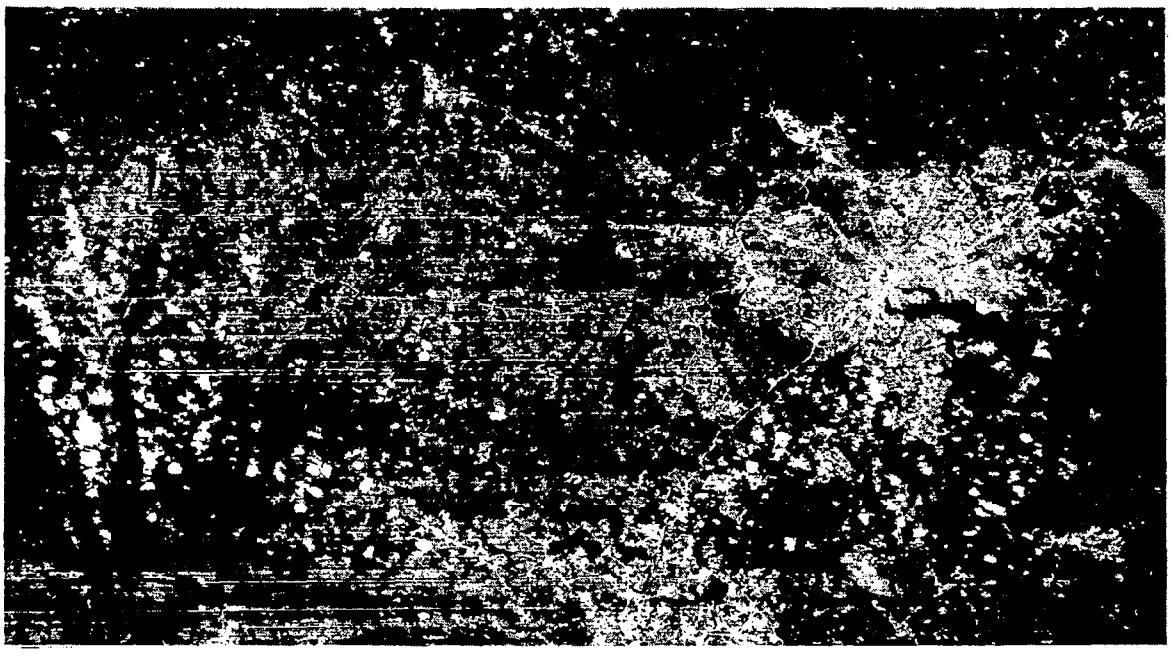


Figure 13B. Frame mean/standard-deviation equalization to Detector 1 on Baltimore area of Scene 2904-14455, Band 4

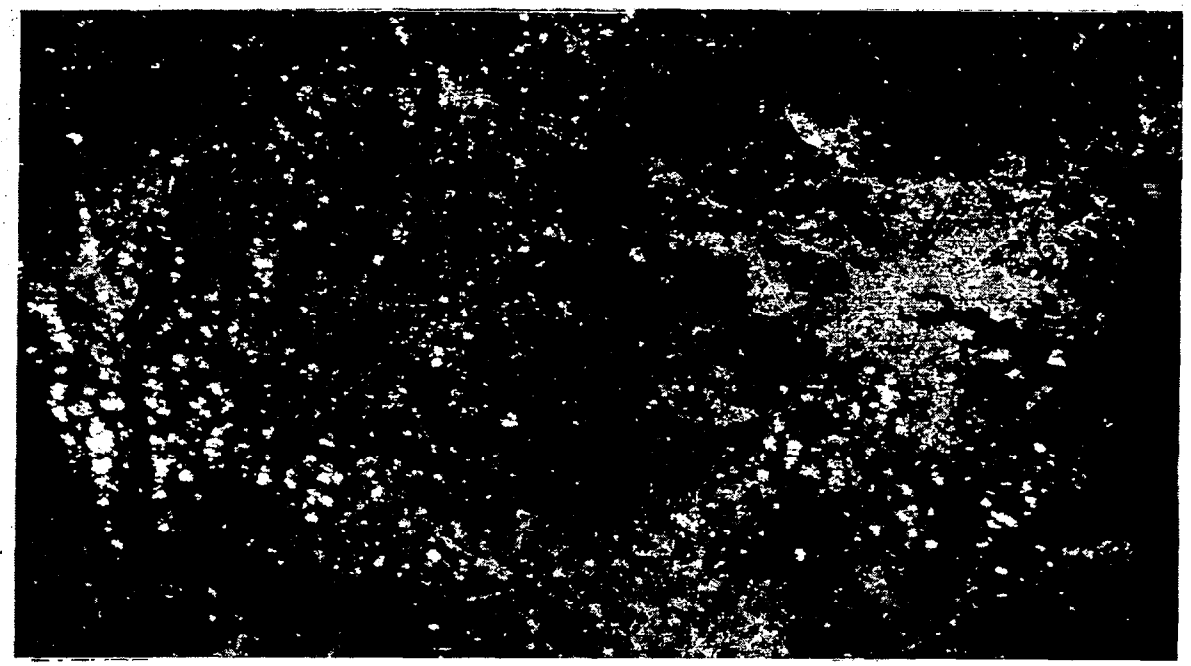


Figure 13C. Sweep mean/standard-deviation equalization on Baltimore area of Scene 2904-14455, Band 4

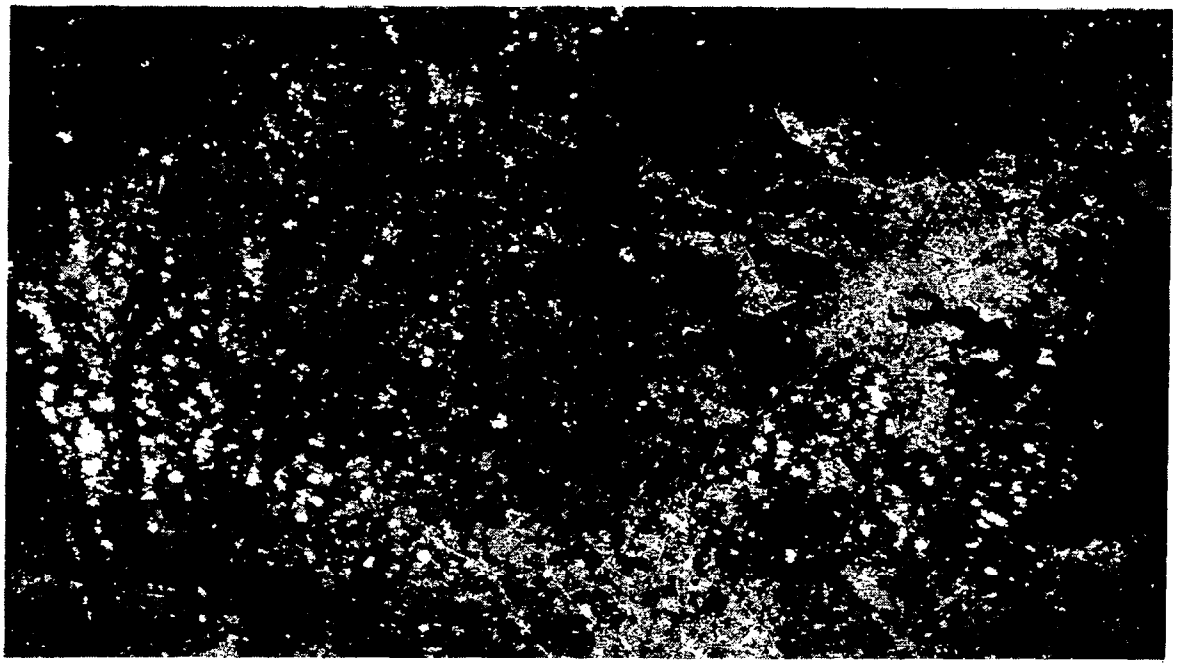


Figure 13D. Local mean/standard-deviation equalization on Baltimore area of Scene 2904-14455, Band 4

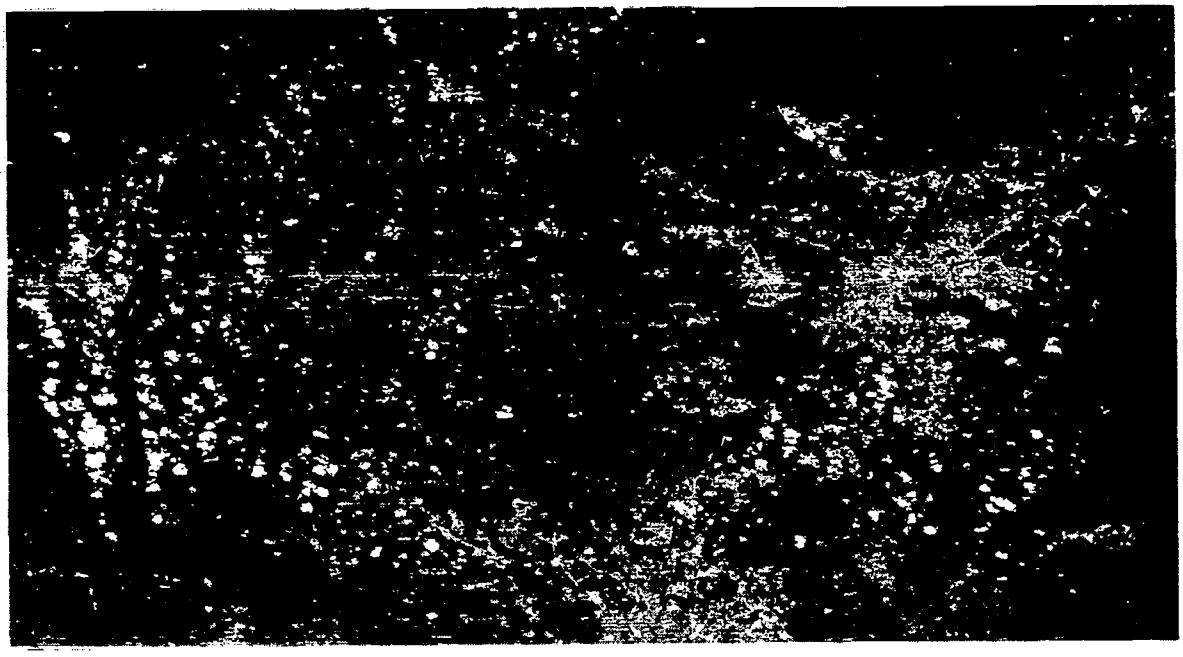


Figure 13E. Sweep histogram equalization on Baltimore area of Scene 2904-14455, Band 4

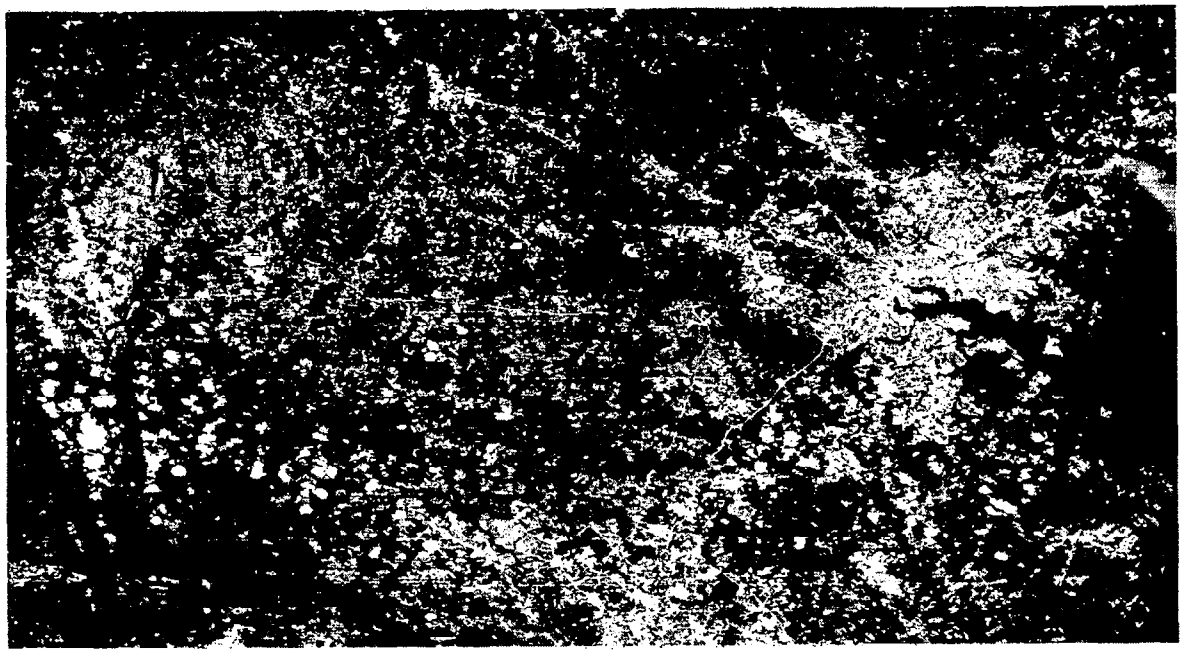


Figure 13F. Local averaging on Baltimore area of Scene 2904-14455, Band 4

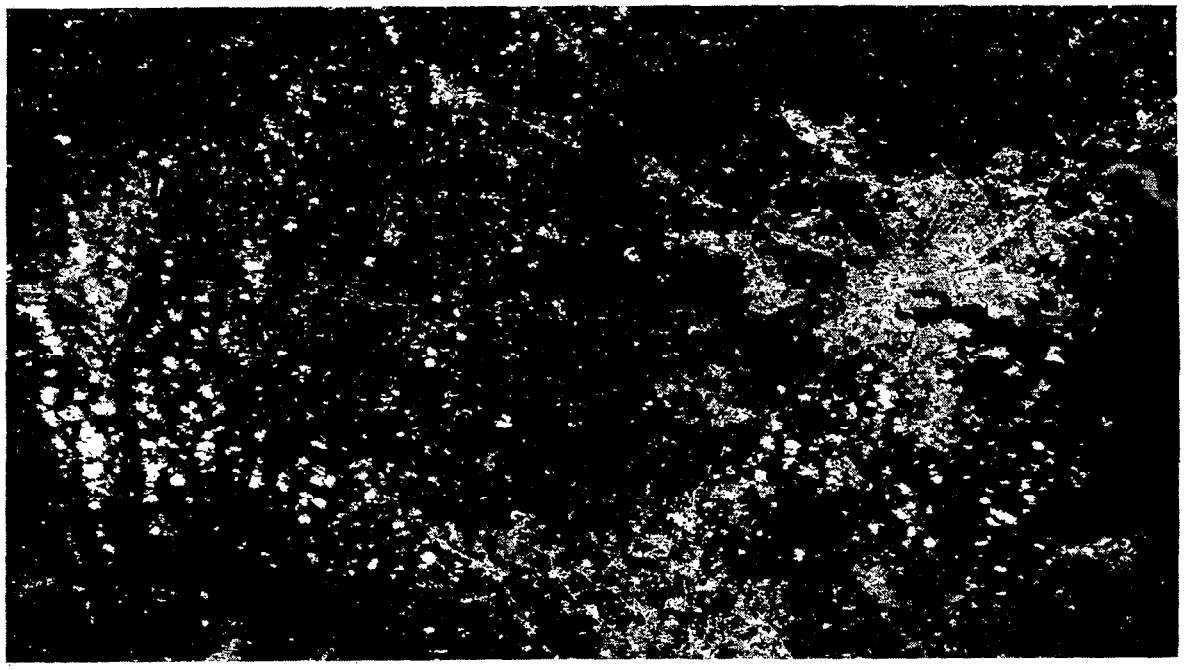


Figure 14. Frame mean/standard-deviation equalization on raw, uncalibrated data from Baltimore area of Scene 2904-14455, Band 4

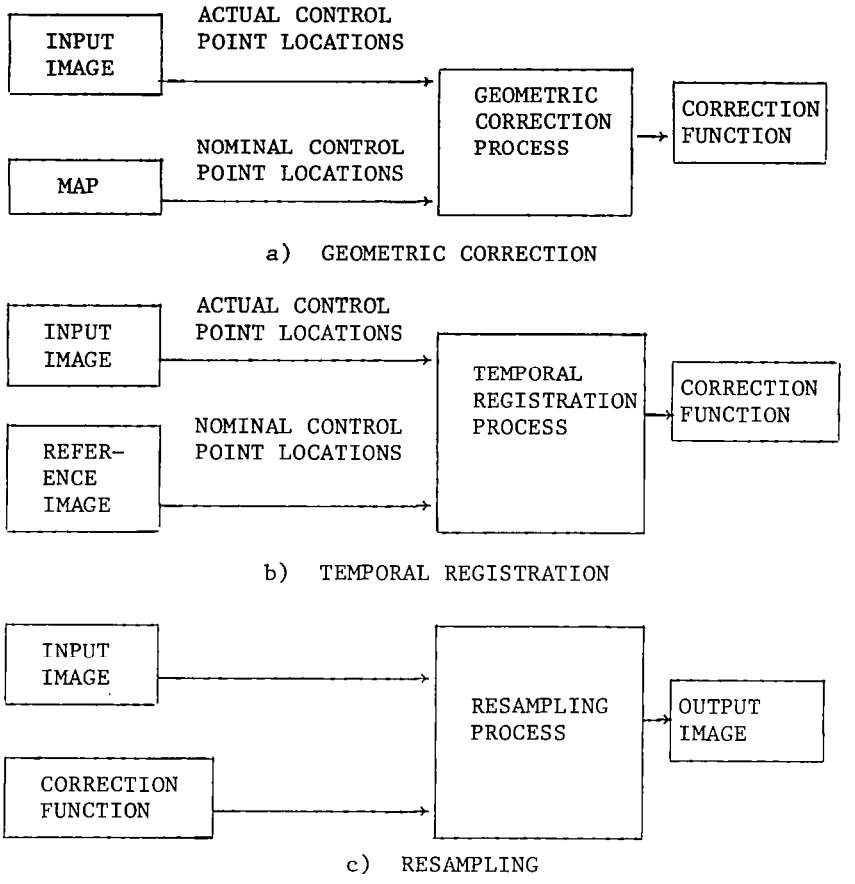


Figure 15. Geometric correction/temporal registration concepts

The steps involved in geometric correction and temporal registration are shown in Figure 16 . In order to meet the required geometric accuracy specifications, control points must be used. Search area processing involves extraction of potential control points from the image data and performing all necessary processing on the extracted data. Control point location is the calculation of pixel locations of the control points in the input data. Error modeling is the computation of coefficients of the geometric error models that will be used to determine the mathematical relationship between the corrected and input images. Output space definition is the characterization of the relationship between the corrected image and a map. Interpolation grid point creation is the final definition of the transformation between corrected and uncorrected images. All of the above steps are completely described below.

Since geometric correction and temporal registration are essentially the same process, they are viewed as a single problem in the following discussion. Wherever they occur, the differences in the two processes are discussed.

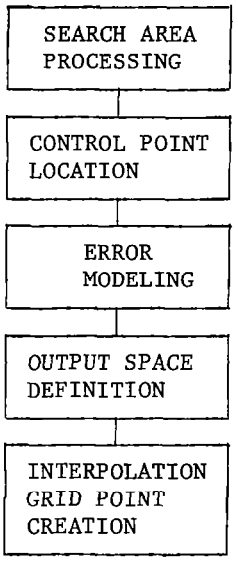


Figure 16 . Geometric correction/temporal registration processing steps

Control Points.

The accuracy of the attitude and ephemeris information available is insufficient to achieve the specified 0.5 pixel geometric correction and 0.3 pixel temporal registration accuracy requirements. However, the errors can be determined and the accuracy requirements can be met through the use of control points in a process described in the following paragraphs. This process was originally developed for the Master Data Processor (MDP) development.

Control points are recognizable features which can be located in the image data through an automatic process called the control location algorithm (CLA). By comparing the actual control point locations with their "nominal" locations, image distortions attributable to the errors in the attitude and ephemeris information are measured and errors are estimated and compensated.

The two types of control points -- geodetic and relative -- differ only in the definitions of their nominal coordinates. Geodetic Control Points (GCPs) are used for geometric correction. i.e., warping an image to a map. Their nominal coordinates are measured from maps and thus GCPs must be features that can be recognized and located

accurately on maps as well as in the image data. This requirement greatly restricts the number of candidate GCPs and prevents many scenes from being geometrically correctable to the desired accuracy.

Relative Control Points (RCPs) are used for temporal registration, i.e., for warping an image to a reference image. Their nominal coordinates are their coordinates in the reference image. Since RCPs must be locatable only in the image data and do not have to appear on maps, an adequate number of them can usually be found for a given scene. Control point characteristics are summarized in Table 6 . In general, a good control point is a feature that is temporally stable both geometrically and radiometrically. It has a correlation function that falls off rapidly from a high peak in all directions. It is sufficiently unique to appear only once in a reasonably small image area (e.g., 200 x 200 pixels).

Table 6 . CONTROL POINT CHARACTERISTICS

Types

- Geodetic
  - Used for geometric correction
  - Accurately locatable in image and on map
  - Nominal coordinates measured from map
- Relative
  - Used for temporal registration
  - Accurately locatable in image
  - Nominal coordinates taken from reference image

Desired Properties

- Geometrically stable
- Radiometrically stable
- Good correlation function
  - High peak
  - Rapid falloff in all directions
- Unique in 200 x 200 pixel neighborhood

Digital representation

- 32 x 32 pixel array (window area)
- Taken from one spectral band

Number required

- Corrections computed for any number ≥ 0

Processing of Control Point Search Areas

The utilization of geodetic or relative control points in the correction of an image begins with the identification and preparation of control point search areas, that is, subimage areas known to contain the features in question. The steps in this process are shown in Figure 17.

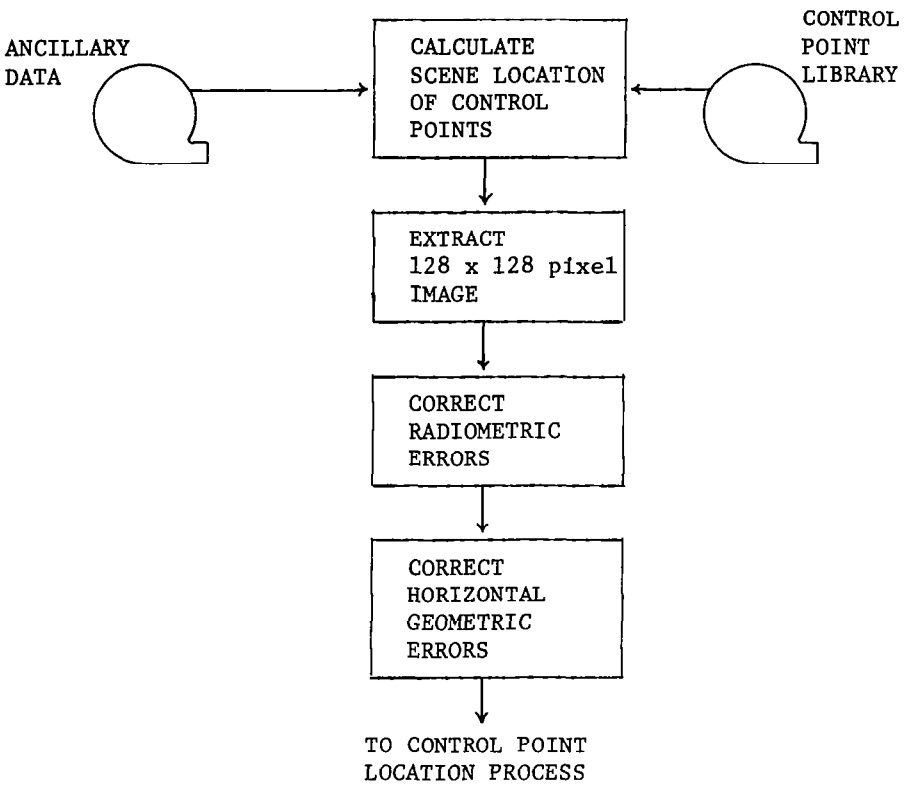


Figure 17. Control point search area preparation

Ancillary data supplied with each scene to be corrected provides the World Reference System (WRS) frame with which the scene is associated and estimates of the spacecraft attitude and ephemeris (altitude and nadir position) during the collection of the scene. The WRS frame identification is used to extract from the control point library the list of control points which should appear in the scene. The control point library information includes the earth-related coordinates of each control point, and these coordinates are used with the attitude and ephemeris estimates to determine the approximate location of each control point in the image data. These locations define the centers of the control point search areas.

Search areas centered at the estimated control point locations are extracted from the image data and are radiometrically corrected with the radiometric correction coefficients computed for the scene. One-dimensional classic four-point cubic convolution is used to remove the local, along-scan errors present in the data (e.g., line length variations and earth rotation) which can all be corrected with the information available at this point in the process. The corrected arrays are then ready for input to the CLA.

Determination of Control Point Image Coordinates

Control points are located in the control point search areas by a CLA developed for the MDP. There were two classes of algorithms and algorithm variants considered for MDP. One class of algorithm was similarity, as represented by the sequential similarity detection algorithm (SSDA). The SSDA essentially computes the sum of the absolute corresponding differences for each of the positions that the reference subimage may take in a search area. A variant of the SSDA is the incorporation of a predictor to indicate, for a particular reference subimage position, that early in the summing further summing may be dropped because this particular position will not be a match. The other class of algorithm considered was two-dimensional cross correlation. For each of the positions that the reference subimage may take in the search area, the normalized sum of products of corresponding elements indicates the degree of matching.

The SSDAs effectiveness and the sensitivity of the predictor are affected by the data. Cross-correlation is a more mathematically robust process. In addition, the Advanced Signal Processor (ASP) used in the MDP implements the Fourier domain version of cross-correlation very efficiently. Therefore, the cross-correlation CLA was chosen for the MDP. At the end of this process, there exists a 97 x 97 cross-correlation surface (see Figure 18).

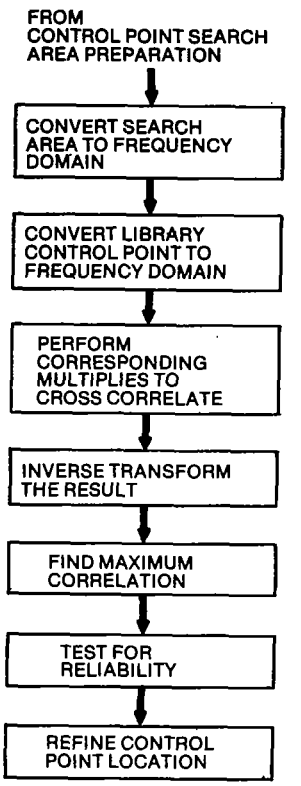


Figure 18a. Control point location process

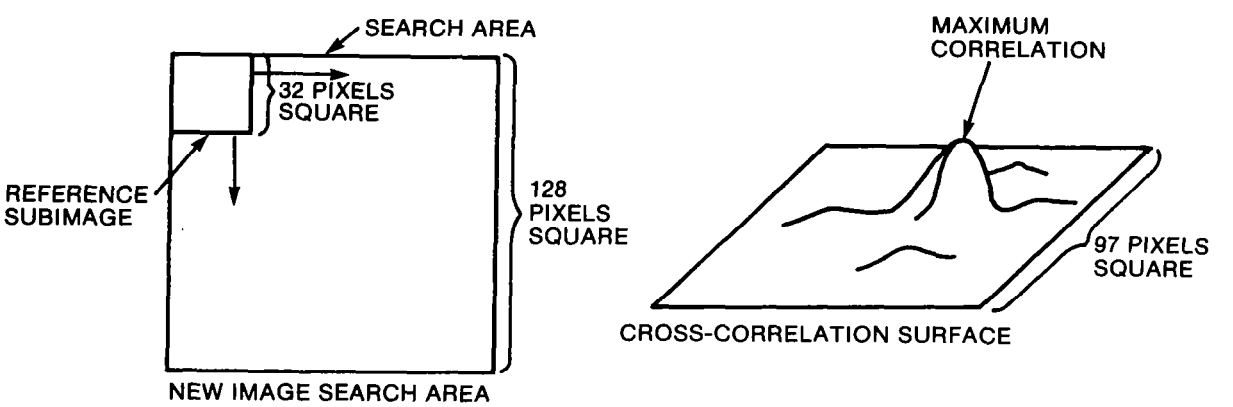


Figure 18b. Search area and cross-correlation surface

The cross-correlation surface is analyzed to determine the location of the maximum correlation sample by comparing all of the values sequentially with a running maximum. An estimate of the exact maximum is obtained by extracting the 5 x 5 matrix of correlation values which has the correlation peak as its (3,3) element, fitting a bivariate, fourth-order surface to this matrix, and computing the location of the peak by Newton's method. The reliability of the peak location is checked by calculating the direction of slowest rolloff or maximum radius of curvature of the 5 x 5 correlation surface from the location of the peak. If the rolloff rate does not exceed a selected threshold, then the control point location is not accepted. Otherwise, the control point location is accepted, and that location can be used for image correction.

Calculation of Geometric Error Modeling Coefficients

Geometric error modeling coefficients are the coefficients of models that describe the distortions present in each image. The coefficients of these models have the same definitions in a geometric correction process as they do in a temporal registration process. Therefore, this topic concerns both types of correction. The error modeling, as shown in Figure 19 is one of the five basic geometric correction/temporal registration processing steps. The resulting coefficients are used in the last two steps of that process.

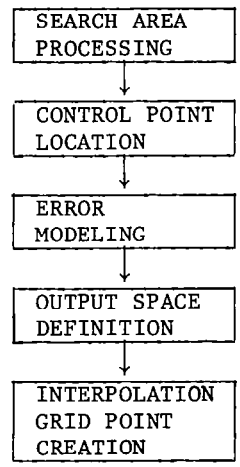


Figure 19 . Geometric correction/temporal registration processing steps

The sources of geometric error can be grouped into three basic classes:

a. Global (low frequency) distortions that can be computed with a priori knowledge without reference to scene data.

b. Global (low frequency) distortions that cannot be known a priori with sufficient accuracy and must be estimated using ground control points found in the scene data.

c. High frequency distortions that must be corrected on a local basis.

The particular errors in each class are shown in Table 7 . The two types of low frequency errors are corrected by the interpolation of grid points, while the high frequency errors are corrected during resampling.

Table 7 . MSS INSTRUMENT ERRORS

Type of Error
Low frequency/GCPs not used
Scan mirror errors
Panoramic distortion errors
Differential scale errors
Scan skew errors
Spacecraft errors
Earth rotation errors
Perspective geometry errors
Map projection errors
Low frequency/GCPs used
Spacecraft attitude measurement errors
Spacecraft ephemeris measurement errors
High frequency
Scan line-length errors
Band-to-band offset errors
Earth rotation errors
Detector sampling delay errors
Mirror velocity errors

The approach used for error modeling is essentially identical to that used by the MDP system. The three steps are shown in Figure 20. To understand the process, it is necessary to understand the coordinate spaces defined in Figure 21. The tangent space represents a cylinder tangent to the earth spheroid at the ground track of the spacecraft nadir.

The first step is to map the observed control point locations to the systematically corrected tangent space. All coefficients of the low frequency error models for which control points are not needed are calculated in this step.

The second step is to calculate the locations of the control points in the fully corrected tangent space. The nominal, earth-fixed coordinates of the control points are used in this step. The only difference between geometric correction and temporal registration occurs here. It consists entirely of the method by which the earth-fixed control point coordinates (used as input to this step) are obtained. For geometric correction, they are based on map measurements; for temporal registration, they are based on a previously processed image.

The third step is to calculate the coefficients of the attitude and altitude error models. This is done by comparing the coordinates of the control points in the two distinct tangent spaces.

For some scenes, there will be no control points available. In this case, the error modeling consists only of calculating the coefficients of the systematic corrections in the first step.

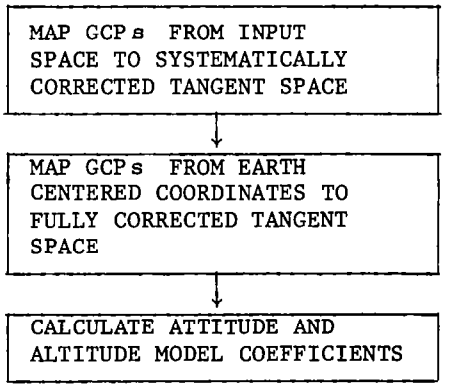


Figure 20. Error modeling steps

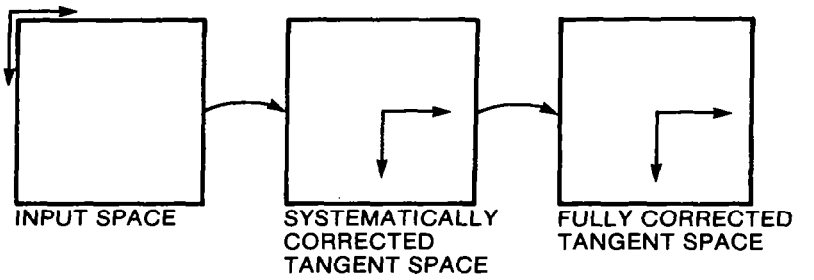


Figure 21. Error modeling coordinate space

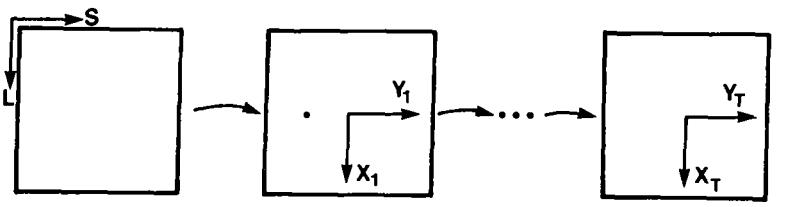


Figure 22. Systematic correction is a series of coordinate transformations

Systematic Correction of Observed Control Point Locations - Systematic correction can be thought of as a series of coordinate transformations that begins with the input space and ends with the systematically corrected tangent space (see Figure 22). Each successive coordinate space has fewer geometric errors than the previous space. The last of these coordinate spaces defines an image in which all low frequency, systematic, geometric errors have been removed.

As shown in Figure 20, systematic correction of the observed locations (in input space) of the control points is the first of the three-error-modeling steps. This step is performed the same way for both geometric correction and temporal registration. The same equations are used in either case.

The steps involved in correcting systematic geometric errors are the following:

1. Convert the input line and sample coordinates to have the coordinate system's origin at the spacecraft nadir.

2. Correct mirror scan distortions and along-scan panoramic distortions. Polynomial models, whose coefficients are determined prior to launch, are used to correct mirror scan distortions. For MSS, a cubic model for along-scan errors is all that is needed.

3. Correct for distortions due to the differential MSS input scales.

4. Correct for along-track MSS scan-skew errors.

5. Correct along-track distortions caused by spacecraft velocity errors.

6. Correct along-scan and along-track errors caused by earth rotation.

7. Correct along-scan errors caused by earth curvature and panoramic projection.

Mapping Nominal Control Point Locations to the Tangent Space - Nominal control point locations are given in earth-centered coordinates. They are input for each control point. These locations are mapped into the fully corrected tangent space. The resulting tangent space locations are the correct locations of those control points, which is why they are called "nominal" control point locations. The algorithm is the one implemented on the MDP system.

As shown in Figure 20, mapping nominal control point locations to the fully corrected tangent space is the second of the three error modeling steps. This step is performed in the same way for both geometric correction and temporal registration. The same equations are used in either case. Nevertheless, it is in this step that the difference between the two types of processing affects the calculations. The method by

which the earth-centered coordinates of the control points were obtained determines which type of correction is being done. If the coordinates were computed from coordinates in a reference image, the control points are relative control points (RCPs), and the image is being temporally registered to match the reference image. If the coordinates were computed from coordinates measured on a map, the control points are geodetic control points (GCPs), and the image is being geometrically corrected to match the map.

The algorithm for mapping points in earth-centered coordinates into tangent space coordinates has the following steps:

1. Calculate the earth-centered coordinates of the spacecraft nadir. This can be calculated directly from the latitude and longitude coordinates of the nadir.

2. Calculate the values of the elements of three rotation matrices. These cause rotations of the coordinate axes such that the resulting axes are parallel to the tangent-space coordinate axes.

3. For each control point, the transformation is evaluated.

Attitude and Altitude Model Coefficients - Attitude and Altitude are quantities that are not known with sufficient accuracy to correct Landsat MSS data. Departures of the spacecraft altitude from nominal, produce scale distortions in the sensor data. For the MSS, the distortion is primarily along-scan and varies with time. Nominally the sensor axis system is maintained so that one axis is normal to the earth's surface and another is aligned with the spacecraft velocity vector. As the sensor departs from this attitude, geometric distortions result. For the MSS, the full attitude time history contributes to the distortion.

The differences between nominal and corrected input-image control point coordinates, along with all applicable ephemeris and AMS data, are processed by a weighted least squares estimator which computes the format center coordinates and the coefficients of the attitude and altitude modules. In these models, spacecraft roll, pitch and yaw variations are described by cubic functions of time (4 x 3 coefficients), and spacecraft altitude variations by a linear function of time (2 coefficients).

The models that will be used are polynomials in time. The form of these polynomials is shown in Table 8. The objective is to determine the coefficients of each of the polynomials.

Table 8. MODELLING POLYNOMIALS

FUNCTION	POLYNOMIALS
Roll	$\phi(t) = \phi_o + \phi_1 t + \phi_2 t^2 + \phi_3 t^3$
Pitch	$\theta(t) = \theta_o + \theta_1 t + \theta_2 t^2 + \theta_3 t^3$
Yaw	$\psi(t) = \psi_o + \psi_1 t + \psi_2 t^2 + \psi_3 t^3$
Altitude error	$\Delta H(t) = h_o + h_1 t$

All of the control point, altitude, and AMS information in the scene being corrected is used by the least-squares estimator. Since the estimator uses all the available data, the models are always the best that can be obtained from the existing information.

Definition of the Output Space

The path of the satellite ground track is not exactly repeatable. It may vary laterally as much as 5000 meters. Therefore, corresponding scenes taken on different days cover only approximately the same ground area. For each nominal Landsat instrument frame, there is a unique, standard reference point. This point, a World Reference System (WRS) format center, is near the physical center of all corresponding scenes (see Figure 23). It is used to define the output spaces and the frame boundaries of the scene. Each WRS format center has a unique heading angle associated with it.

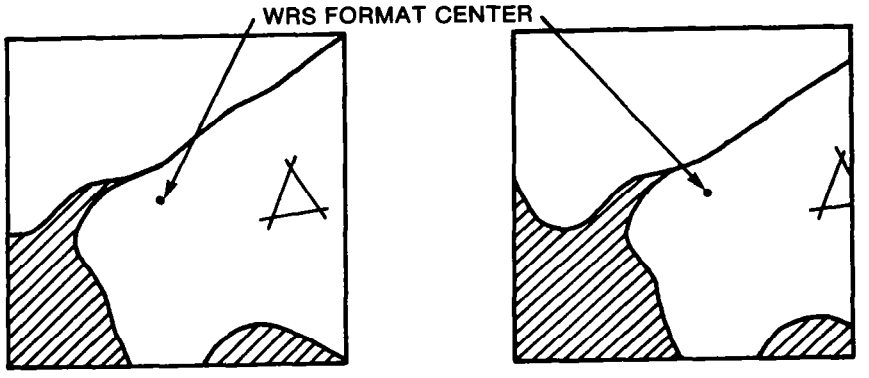


Figure 23. WRS format centers in two different scenes

The output space is defined by the following conditions:

a. The WRS format center falls on the middle line of the output frame.

b. The WRS format center is displaced from the center pixel by an integral number of pixels.

c. The angle between a parallel of latitude at the WRS format center and the negative vertical axis is the WRS heading angle (see Figure 24).

d. Each output-image pixel represents a square region of a fixed, constant size on the ground.

With these conditions, the fixed set of WRS format centers (there will be more than 60,000 of them worldwide) controls the framing of image data.

The reasons for this use of the WRS format center are to have the image data approximately centered in the output frame (thereby minimizing the amount of data that constitutes an output frame) and to have nominal registration of all scenes of the same ground area (see Figure 25 b).

The resulting pixel coordinates of the WRS format center, the unique geodetic coordinates of the WRS format center, the unique WRS heading angle, and the output pixel size characterize the output frame. The WRS format center occupies one point on a theoretically infinite sampling lattice. Figure 26 shows the relationship between the various "format centers" on a pixel-by-pixel basis.

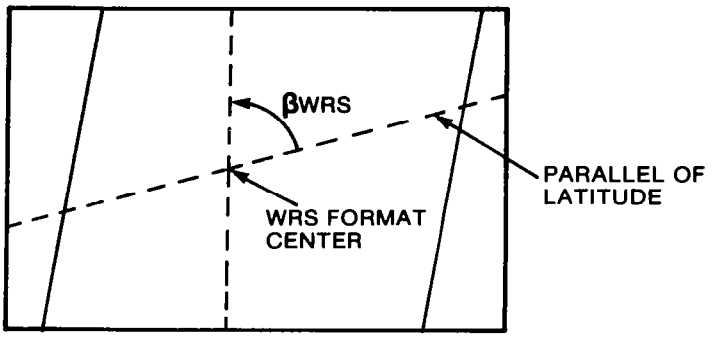


Figure 24. WRS heading angle

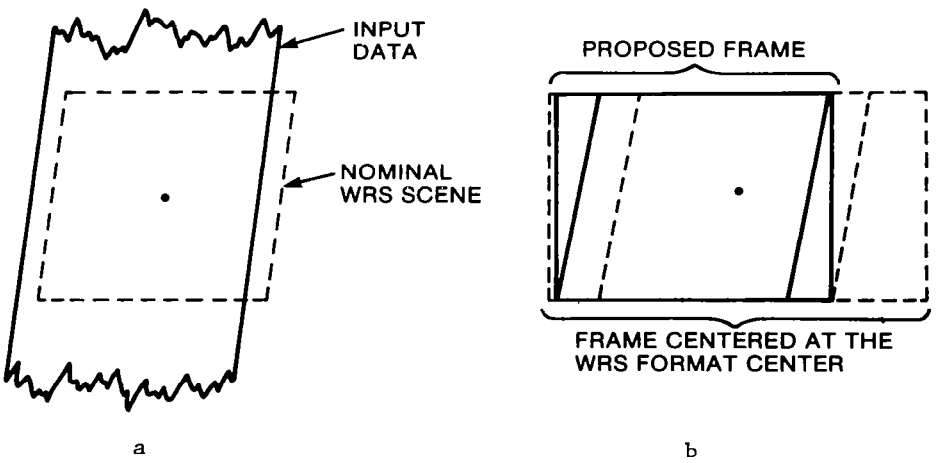


Figure 25. Proposed framing minimizes width of output frame

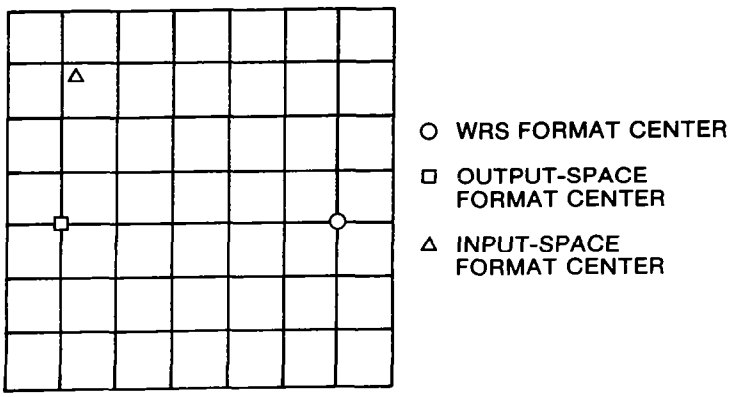


Figure 26. Locations of "Format Centers" in output space

Interpolation Grid Point Correspondence

To create a corrected image, it is necessary to know the relationship between the locations of pixels in the input space and the output space. This relationship takes the form of a function that maps the output space into the input space. For each output pixel, the corresponding input space location at which the data will be resampled must be computed. In order to minimize the computations involved, only selected grid points in the output space are mapped to the input space by using the error modeling equations. The final output space to input space mapping is an interpolation between the selected grid points.

The positions of grid points can be controlled in only one of the two spaces involved. In order to make the resampling process efficient, the grid points are defined on a rectangular lattice in the output space. Any other distribution of the grid points would make resampling very unwieldy or impossible. Figure 27 illustrates this primary grid point correspondence. The output space locations of the primary grid points are somewhat arbitrary and are fixed. In the space-to-space mapping process, the corresponding input space locations of the primary grid points are determined.

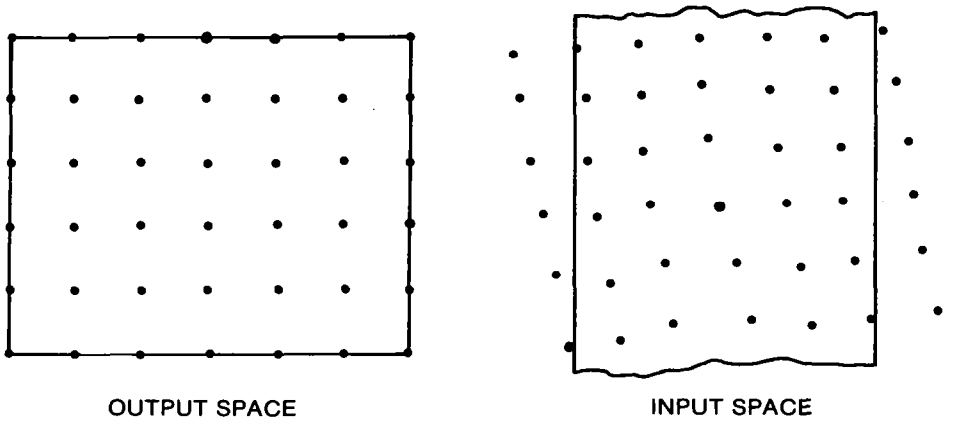


Figure 27. Primary grid point correspondence

Space-To-Space Mapping Approach and Tradeoffs - Space-to-space mapping is the process by which a known location in the output space is mapped through geometric distortion models to the corresponding input space location. A rectangular array of primary grid point locations in output space is mapped into the input space, as shown in Figure 27 . The resulting output from the space-to-space mapping is a table of corresponding output space and input space grid points.

The geometric distortions in scanner data lend themselves to a correction transformation that flows from input space to tangent space (an "input-to-output" mapping). Because of the timing relations involved, it is very difficult to obtain an output-to-input mapping with an accuracy equivalent to that of the corresponding input-to-ouput mapping. The attitude, altitude, and earth rotation corrections are all functions of time. The correct time at which a pixel was imaged can easily be determined in input space. In tangent space, that time can only be approximated. Therefore, an output-to-input mapping is usually less accurate than an input-to-output mapping.

From the standpoint of accuracy, it is desirable to use an input-to-output mapping to determine the locations of the primary grid points. An iterative technique using accurate error modeling coefficients is used to obtain the input space primary grid point locations. This is a variation of the technique used on the MDP system.

The coordinate spaces and functions involved are shown in Figure 28. Each primary grid point (X,Y) in the UTM output space is mapped to the point $(\overline{X},\overline{Y}) = L_T(X,Y)$ in the tangent space. The objective is to find the point (U,V) in input space such that $I_T(U,V) = (\overline{X},\overline{Y})$. An initial approximation to (U,V) is the point (U_1,V_1) obtained from the approximate error-modeling function T_I. Then iteration using the function I_T is performed until convergence is obtained.

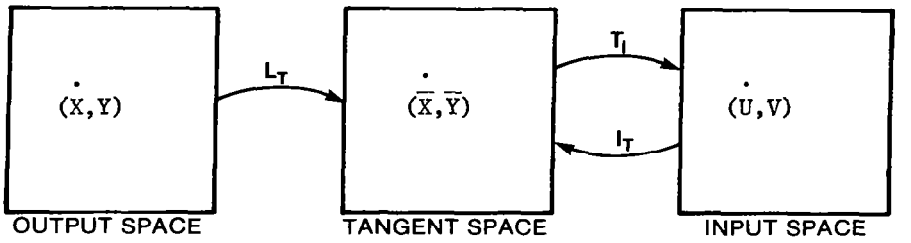


Figure 28. Iterative technique for space-to-space mapping

Details of Space-to-Space Mapping - The fixed array of primary grid points is mapped from the UTM output space into tangent space using the map projection function L_T, as shown in Figure 28. This results in an array of tangent space points for which corresponding input space points must be found.

A set of approximate error modeling equations is used to get initial estimates of the input space locations of the primary grid points. This mapping uses equations that represent inverses of the accurate mapping equations described above. The following distortions are corrected:

a. Attitude and altitude
b. Earth curvature and panoramic projection
c. Earth rotation
d. Spacecraft velocity
e. Scan skew

f. Differential input scale

g. Mirror velocity

It should be noted that the time is computed from the tangent space coordinates and is only approximately correct. This affects both the attitude/altitude correction and the earth rotation correction.

Once the estimated input space locations of the primary grid points are obtained from the function T_I, an iterative procedure is used to obtain the correct input space location. Let (X,Y) be a tangent-space location, and let (U_i,V_i) be the corresponding estimated input space location. The accurate error-modeling function I_T is evaluated at (U_iV_i) to get a tangent space point (X_iY_i). The process is iterated by incrementing the coordinates (U_iV_i) by the difference (X Y) $-(X_i,Y_i)$ and repeating the process until that difference becomes small. MDP experience has shown that the process will converge to an acceptable error within three iterations. (See Figure 29).

Error Analysis

It is frequently desirable to know the geometric accuracy of a corrected Landsat MSS image. In this section, two statistical methods of assessing the accuracy are discussed. Both methods have been used satisfactorily in the past (e.g., MDP, ERTS System Studies).

Comparison with a Map - One way to measure image accuracy is to compare precisely the image to a set of very accurate maps (such as the 7½-minute USGS Topographic Maps). This involves locating a set of geodetic check control points (GCCPs) in the corrected image and on the maps. These may include some of the GCPs used to correct the scene. They must also include clusters of points to allow the estimation of GCCP measurement errors. An analysis of variance technique can be used to compute the image relative geometric accuracy.

In determining the error due to the geometric correction (or temporal registration) process, differences (measured in meters) between the observed (image) and nominal (map) values of the coordinates of corresponding features are formed in the horizontal and vertical directions. These differences are called residual errors. A given residual error is the sum of the error due to the geometric correction process (the image error) and the error due to the process of locating the feature both in the image and on a map (the measurement error). Over a cluster of closely located features, the image errors are nearly constant while the measurement errors will exhibit the variability associated with the normal probability distribution. If features which can be grouped into clusters can be selected, the observed residual error can be expressed as:

$$y_{ij} = a_i + b_{ij} \text{ for } j = 1,\ldots,n_i \text{ and } i = 1,\ldots,A$$

where

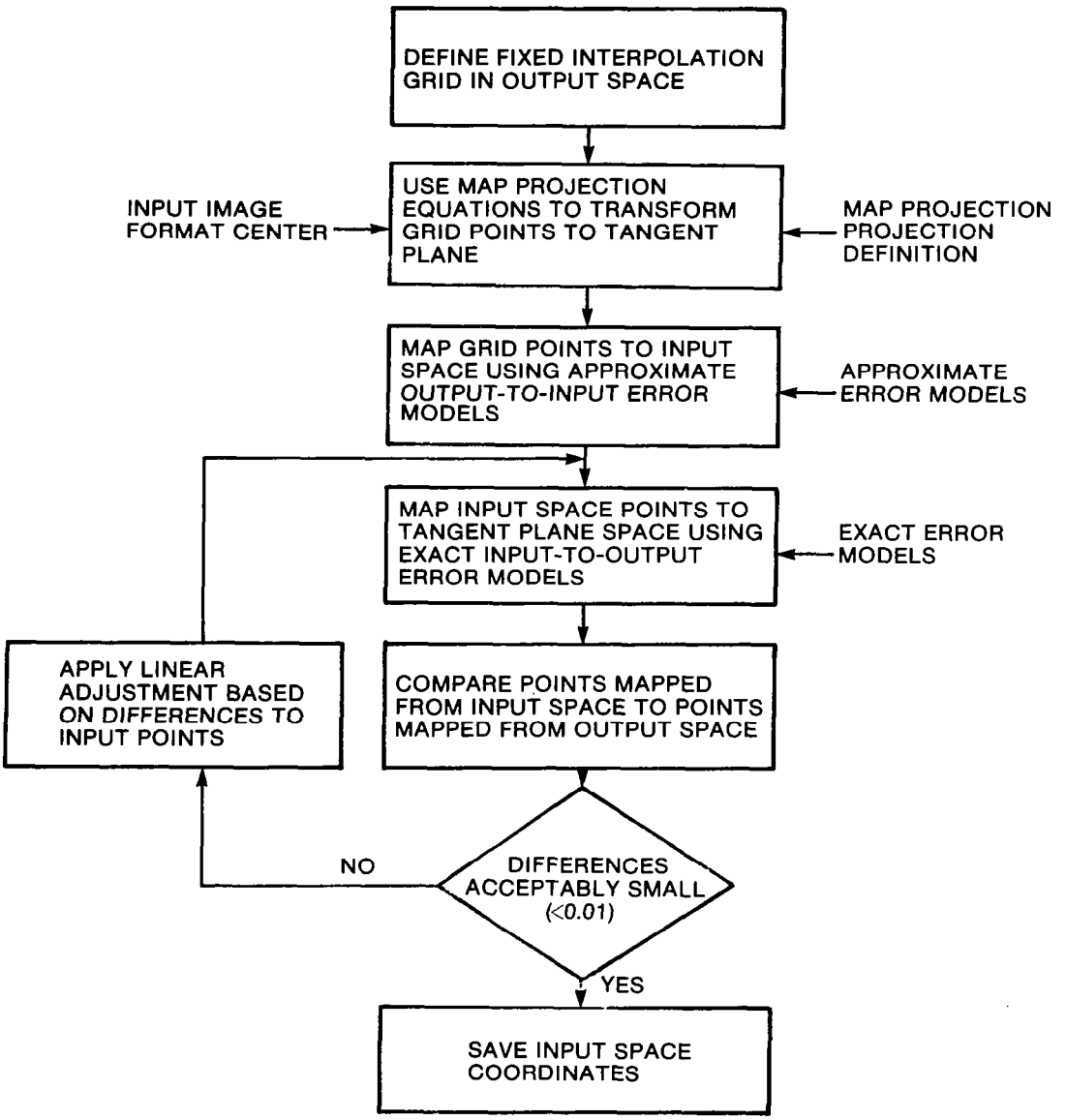


Figure 29. Space-to-space mapping

y_{ij} = the residual error of the j^{th} feature in the i^{th} cluster

b_{ij} = the measurement error

a_i = the image error

A = the number of clusters

n_i = the number of elements in the i^{th} cluster.

A modification of the classical analysis of variance can then be used to estimate the variance of the image errors and the variance of the measurement errors.

Automatic Covariance Analysis - A second estimate of image accuracy may be obtained from the covariance matrix of the least squares attitude/altitude fit used in the geometric correction error modeling. The method requires no further measurement of control point locations (as in the previous case), and the results can be calculated as part of the correction process.

In the geometric correction process, all known systematic geometric errors are removed first. The remaining errors are corrected by the attitude/altitude model, and it is assumed that a correctly located set of GCPs will result in an image with no error. Therefore, errors in GCP locations are the ultimate cause of image error. It is assumed that these sources of error are normally distributed with mean zero and variance known from past experience. There is a linear relationship between the errors in GCP locations and resulting errors in the coefficients of the attitude/altitude model. Therefore, the errors in the model coefficients will also be normally distributed, as will the image errors after application of the mapping function incorporating the attitude/altitude model.

A given set of GCPs will lead to a certain image accuracy which will depend on the distribution of the GCPs and the accuracy with which the difference between the map location and image location of each GCP can be measured. Using the above assumptions, the image accuracy from the given set of GCPs may be estimated. The method of estimating the image accuracy is the standard one of determining the effects of random errors in the data used to make the corrections by mapping the covariance matrix of the attitude/altitude model coefficients (obtained in the process of determining these coefficients) onto a set of points distributed over the image plane, and by similarly taking into account the errors due to grid point location (in the mapping process) and mirror velocity profile, which can be modeled as random processes.

Mapping of the attitude/altitude covariance matrix (plus the effects of grid point location and mirror model) to a point in the image plane leads to an error ellipse at that point which, because of the nature of the observation matrix, has its principal axes aligned with the horizontal and vertical directions. The probability that the true image point falls within the ellipse centered at the mapped point whose principal axes are

multiplied by k is

$$P(2,k) = 1 - e^{-k^2/2},$$

so that the 90% confidence ellipse can be found by multiplying the principal axes of the computed error ellipse by 2.15, and the 95% ellipse by multiplying by 2.45.

Resampling Methods

One of the requirements normally imposed upon image processing when the image is in digital form is that the output pixel lattice be regular, or equispaced, in the output space. This is needed for two reasons: compact data storage, and the limitations of many film recorders. If the output pixel lattice is not regular, then additional information must be carried to specify pixel location. The last requirement precludes the simple repositioning of the original image pixels. The only way to change geometry and specify the target pixel location, too, is to "resample" the original image. Resampling here consists of calculating the location of a particular target pixel in the original image and interpolating over the surrounding original pixels to find out the output intensity. During processing, the target pixels are considered output and the original pixels are considered input. Figure 30 shows the relationship between input and output pixels.

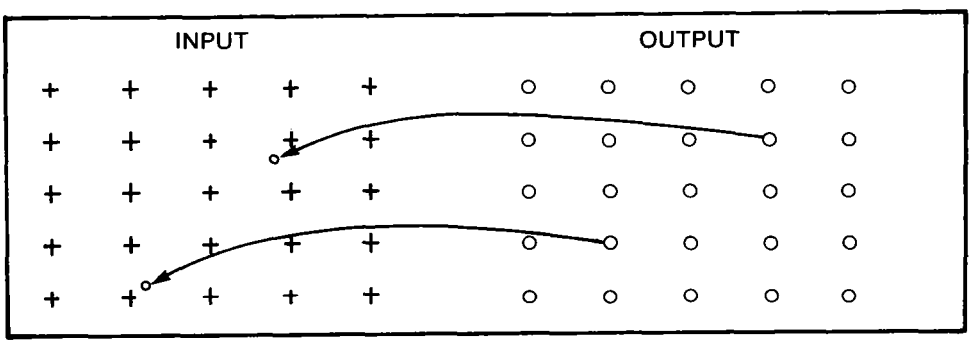


Figure 30. Locating output points in input space

In the following discussion, the calculation of the output pixel location in the input image is presumed, and therefore, the discussion focuses on the different interpolators. Two types of interpolators are specifically considered: nearest neighbor resampling, and cubic convolution resampling.

Resampling Algorithms

Nearest Neighbor - Nearest neighbor is a single, specific method of determining the intensity value of an output pixel located in the input image. The intensity value of the input pixel that is closest to the output pixel is selected as the output pixel intensity (see Figure 31).

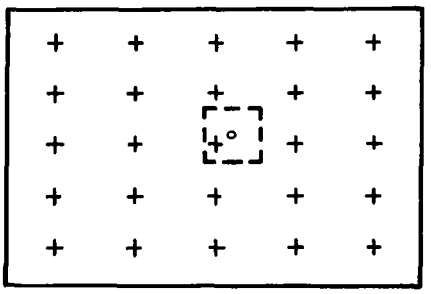


Figure 31. Nearest neighbor resampling

Cubic Convolution - Cubic convolution is a family of resamplers that approximate the SINC, or (SIN x)/x function. The SINC function is the theoretical perfect resampler, but requires an infinite number of terms. Cubic convolution substitutes a truncated piecewise cubic approximation to the SINC function, so that the amount of processing is feasible for large image data sets. Cubic convolution is a one-dimensional process that must be repeated to provide for two-dimensional resampling. This is illustrated in Figure 32, where four horizontal resamplings provide the values for a final vertical resampling. The final vertical resampling then provides the output intensity. This is for a so-called four-point resampler; a six-point resampler would require seven resamplings.

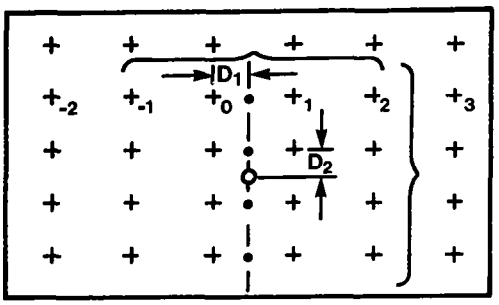


Figure 32. Two-way cubic convolution resampling

Four members of the cubic convolution family have been analysed to see their advantages and disadvantages:

a. Classic four-point cubic
b. Optimized four-point cubic
c. Six-point cubic
d. Optimized six-point cubic

The "optimized" designation refers to a measure of smoothness that was optimized for two cases. This smoothness measure is the reduction of the integral of the second derivative of the resampling kernel. The equations of the four interpolating functions are shown in Table 9.

Table 9. RESAMPLING ALGORITHMS

Type	Equation
Classic four-point cubic	$I_{OUT} = D\{D[D(-I_{-1}+I_0-I_1+I_2)+(2I_{-1}-2I_0+I_1-I_2)] + (I_1-I_{-1})\} + I_0$
Optimized four-point cubic	$I_{OUT} = \frac{D}{2}\{\frac{D}{2}[\frac{D}{2}(-I_{-1}+3I_0-3I_1+I_2)+(2I_{-1}-3I_0+2I_1-I_2)] + (I_1-I_{-1})\} + I_0$
Six-point cubic	$I_{OUT} = \frac{D}{2}\{\frac{D}{2}[\frac{D}{2}(I_{-2}-I_{-1}+2I_0-2I_1+I_2-I_3) +(-2I_{-2}+3I_{-1}-4I_0+2I_1-I_2+I_3)] + (I_{-2}-2I_{-1}+2I_1)\} + I_0$
Optimized six-point cubic	$I_{OUT} = \frac{D}{5}\{\frac{D}{5}[\frac{D}{5}(I_{-2}-3I_{-1}+6I_0-6I_1+3I_2-I_3) +(-2I_{-2}+7I_{-1}-11I_0+7I_1-2I_2+I_3)] + (I_{-2}-4I_{-1}+4I_1-I_2)\} + I_0$

The four versions of cubic convolution resampling were compared and ranked with respect to four qualities: step response overshoot, interpolation accuracy, image appearance, and computer running time. The rankings are summarized in Table 10.

Table 10. RELATIVE RANKINGS OF CUBIC CONVOLUTION ALGORITHMS

TYPE	Step-Response Overshoot	Reconstruction Error	Image Appearance	Running Time
Optimized four-point cubic	1	2	4	2
Optimized six-point cubic	2	1	3	4
Six-point cubic	3	2	2	3
Classic four-point cubic	4	2	1	1

Step response shows the resampler's ability to reproduce rapidly changing image patterns and the amount of overshoot that results. Overshoot affects both data fidelity and image appearance. It is overshoot that causes extra borders that parallel sharp transitions in images. Examining Table 10 shows that the resampling algorithms with higher data fidelity produced lower-ranked image appearance, and conversely. It might be inferred that the distortions introduced by those resamplers ranked high in image appearance emphasize edges and give an appearance of greater definition.

The classic four-point cubic convolution is a good compromise between data fidelity and image appearance and is the most efficient in terms of inner loop processing cycles. The algorithm is implemented on the Master Data Processor (MDP).

High-Frequency Distortions

There are several geometric errors present in raw MSS data that vary from one scan line to the next. These errors are the following:

1. Line length variations.
2. Earth rotation errors.
3. Sampling delay errors.
4. Band-to-band offsets.
5. Mirror velocity errors.

All of these errors represent displacements in the along-scan direction. Since these distortions vary on a line-by-line basis, they are called high-frequency distortions.

Since the high-frequency geometric errors are local in nature, they are not removed by applying the global transformation described in the Geometric Considerations section. The situation is as shown in Figure 33 . The global geometric transformation G maps locations in the corrected output space to locations in the high-frequency-corrected input space. The local transformation L maps the high-frequency-corrected input space into the raw input space.

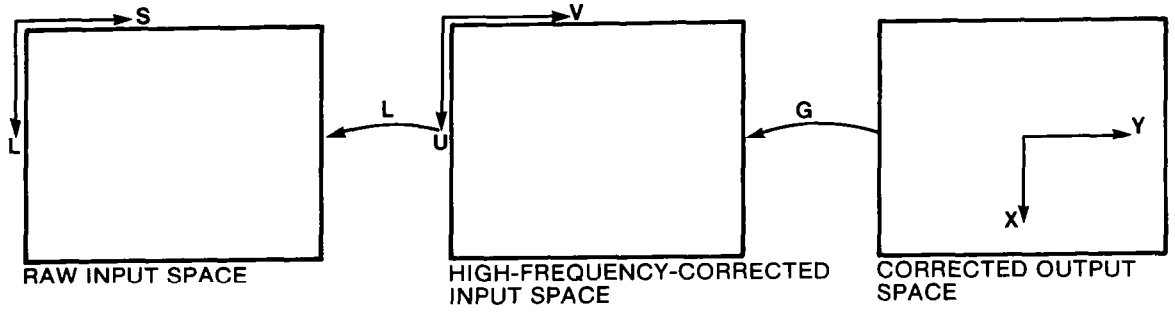


Figure 33. Geometric correction and resampling spaces.

It is important to understand that the high-frequency-corrected image is not explicitly created during the resampling process. The functions G and L are evaluated during the resampling step to obtain locations in raw input space. The raw data values are then used to compute the corresponding corrected output-space values. It would be possible to create a high-frequency-corrected image by performing a one-dimensional horizontal resampling of the raw input data using a cubic convolution algorithm. Such an image could then be resampled using only the global transformation G. In this case the data would have gone through two resampling algorithms. Since it is preferable to degrade the data as little as possible, the high frequency correction should be implemented as part of the basic resampling process.

Implementation Strategies for Resampling

Three different ways to implement the resampling process on a digital computer will be discussed in this section. Then a comparison of these algorithms will be presented.

Resampling via Hybrid Space - Resampling is performed with one-dimensional algorithms that are used separately in two dimensions. Conceptually, the resampling can be viewed as the creation of two distinct resampled images: one that has been resampled horizontally only, and one that has been resampled both horizontally and vertically. The input space is resampled horizontally to form an intermediate image called hybrid space. Hybrid space is then resampled vertically to form the final output space. This is shown in Figure 34.

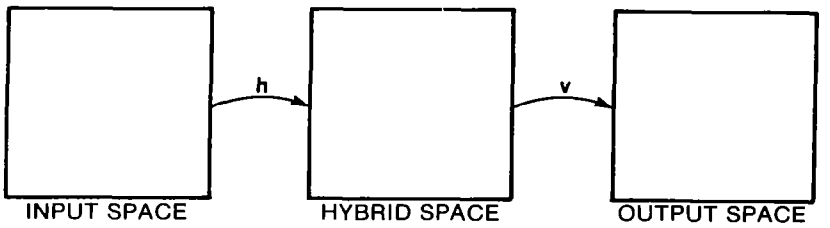


Figure 34. Resampling spaces

The locations at which horizontal and vertical resampling take place are illustrated in Figure 35 . The hybrid space pixels are the intersections of output space columns with input space rows. Horizontal resampling is performed on input space rows, and vertical resampling is performed on hybrid space columns. In this way, cubic convolution can be performed with the minimum of calculations.

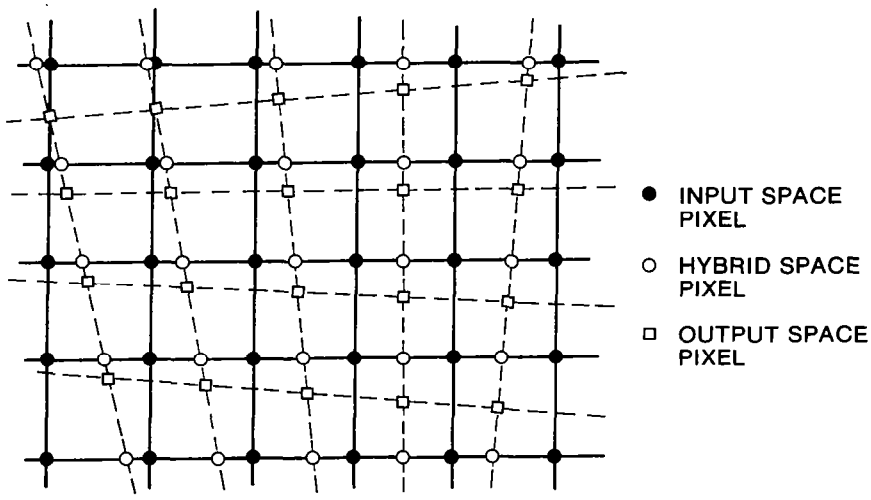


Figure 35. Horizontal and vertical resampling

Construction of the hybrid space requires a second set of grid points, and construction of the output space requires a third set of grid points. These interpolation grid points are derived by cubic interpolation methods from the primary grid points.

High frequency geometric errors caused by scan line length variations, band-to-band offsets, earth rotation, and detector sampling delay are removed at resampling time. This is done during horizontal resampling.

The use of interpolation grid points to speed the resampling process is a technique that was perfected for the MDP system. It permits high speed resampling and maximal system throughput.

Resampling Directly - An alternate way to view the resampling process is to consider it as a transformation that has no intermediate hybrid space, as shown in Figure 36. Resampling is still performed with one-dimensional algorithms that are used separately in

two dimensions. However, all of the calculations for an output pixel value are completed before those for the next output pixel value are begun. This is shown in Figure 37.

The difference between resampling via hybrid space and resampling directly can be seen by comparing Figures 35 and 37. The hybrid-space pixels in Figure 35 are comparable to the intermediate values in Figure 37. When resampling directly a new set of intermediate values is computed for each output space pixel. When resampling via hybrid space, each hybrid-space pixel value is used to calculate four output-space pixels.

Point Shift Algorithm - The point shift algorithm is a method of performing the nearest neighbor resampling algorithm on a computer. It is particularly suited to images in which the differences in scale and rotation between the output and input images are small. This is illustrated in Figure 38, which shows the correspondence between a segment of an output-space line and input space. In this example, nine consecutive input pixels from one line are the nearest neighbors of nine consecutive output pixel from one line. In general, each output line of a nearest neighbor resampled image will be composed of segments of lines from the input image. It is possible to determine the lengths of these segments from the partial derivatives of the transformation between the output and input spaces. Then the output image array may be constructed by transferring each of these segments of the input image array with a single computer instruction.

Figure 36. Resampling spaces

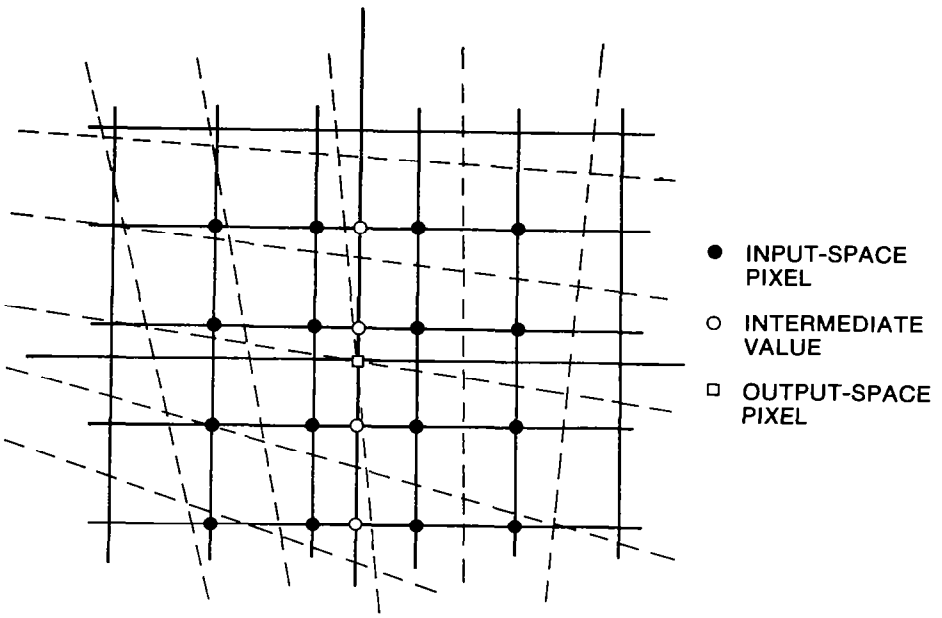


Figure 37. Resampling directly

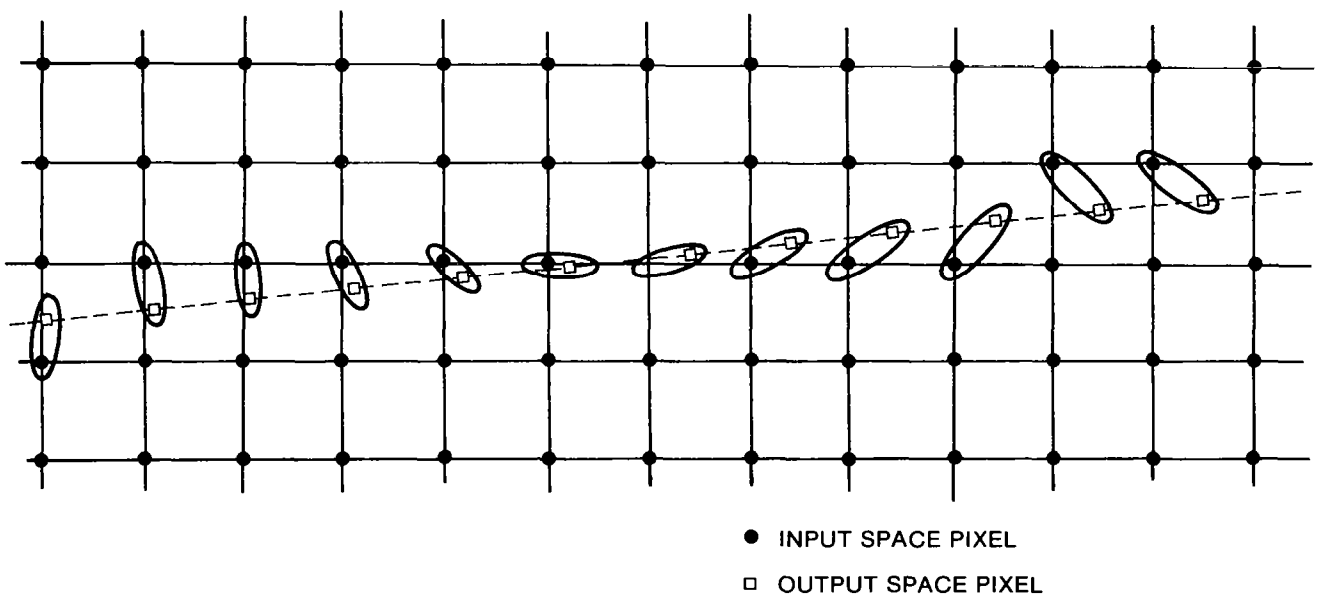


Figure 38. The point shift algorithm

Comparison of Resampling Implementation Strategies

The three resampling strategies can be compared with respect to several criteria. A summary of the rankings of the strategies is given in Table 11. The implementation strategies are rated independently for nearest neighbor and cubic convolution resampling. While Table 11 may seem to indicate that direct resampling is the best strategy, that is not necessarily the case. Some considerations that may apply are discussed below.

Resampling via hybrid space requires considerably less computation than does resampling directly. For a cubic convolution algorithm, using hybrid space requires only two evaluations of the one-dimensional algorithm per pixel, while resampling directly requires five evaluations per pixel. Even though the use of hybrid space requires more instructions to control the flow of data than are necessary for direct resampling, it is still a considerably faster algorithm. It is expected to require no more than half of the computer time needed for direct resampling. It is conceivable that the speed factor alone could make hybrid-space resampling the best choice. It depends on how much data will be resampled by the program.

Speed is much less important a factor in nearest neighbor resampling than it is in cubic convolution resampling. This is because the computer running times are much less for nearest neighbor.

Direct resampling and hybrid-space resampling can use the same software framework for different resampling algorithms. (This is not true for point shift, which is a nearest neighbor technique.) However, in the case of nearest neighbor resampling, the use of hybrid space can introduce some geometric error. Figure 39 shows how this can happen.

In this example, the slope of the left output space column is such that neither of the two hybrid space pixels is the correct nearest neighbor for the output space pixel.

In summary, it appears that if running time is the most important factor, hybrid space for cubic convolution is the proper choice. For nearest neighbor, the best choice is probably direct resampling. However, if hybrid space or direct resampling was chosen for cubic convolution, the same technique should probably be used for nearest neighbor.

Comparison of Interpolation and Deconvolution Based Resampling Algorithms

There are two basic approaches to resampling a digital image. One is to apply any one of a large number of interpolation algorithms which use polynomial, geometric or other functions to generate data points at new locations in a grid. These methods operate on the frequency content of the sampled image and produce a new image which is a filtered replica with the filter characteristic being a function of the algorithm chosen. Deconvolution based algorithms on the other hand are designed to compensate for the filtering effects of a particular scanner aperture and are thus sensor dependent. A deconvolution algorithm designed to compensate for the Landsat scanner instantaneous field

Table 11. COMPARISON OF RESAMPLING STRATEGIES

	NEAREST NEIGHBOR				CUBIC CONVOLUTION		
	HYBRID SPACE	DIRECT	POINT SHIFT		HYBRID SPACE	DIRECT	POINT SHIFT
SPEED	3	2	1		1	2	N/A
SCALE DIFFERENCES	2	1	3		2	1	N/A
ROTATION	2	1	3		2	1	N/A
GEOMETRIC ACCURACY	2	1	1		1	1	N/A
SOFTWARE COMPLEXITY	2	1	3		2	1	N/A
SOFTWARE FLEXIBILITY	2	1	3		2	1	N/A
DEVELOPMENT COST	2	1	2		2	1	N/A

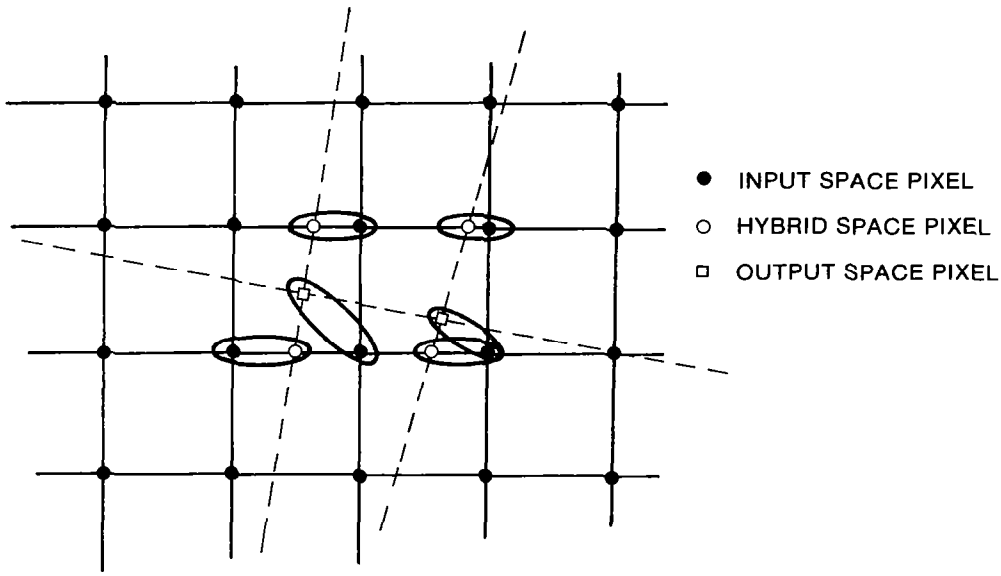


Figure 39. Geometric errors in hybrid space nearest neighbor resampling

of view and at the same time interpolate to generate new points was tested on the data used in this study. The algorithm in effect produces a smooth surface as does cubic interpolation while increasing high frequency response in a controlled manner.

A comparison test was carried out to enable a visual evaluation of the deconvolution algorithm. Three images were generated of an area from the Landsat data east of Salisbury, Maryland including the municipal airport for that city. The first image was generated by duplicating Landsat pixels to achieve approximately a 20 meter sample spacing. This result is shown in Figure 40a. Note the blocky appearance of the image. Next a standard cubic interpolation process was employed to generate new data points to achieve the same sample rate. This output is shown in Figure 40b. Note the smooth but blurred appearance of the image. The deconvolution and interpolation algorithm was then applied to the data and the result in Figure 40c was produced. Note that the image is smooth but has more variation, structure and sharpness than the other images. The result is presented for evaluation by the user and if the deconvolution based result is judged desirable then the capability is available to carry out the processing. The effect of these processes on classification performance has been shown to be unpredictable with some cases showing improvement and others not (Reference 2). Such effects are beyond the scope of this report.

Landsat Image Filtering for Enhancement

It is possible to modify Landsat digital data in such a way as to enhance certain aspects of the image. Two such aspects are contrast and edge definition. The purpose of both types of enhancement is to make certain features more visible in a photographic rendition of an image. An enhanced image usually represents a less accurate reconstruction of the true data than the original Landsat digital radiometric data. However, enhanced images generally appear to be more pleasing to view and to contain more information.

Contrast Enhancement

The usual purpose of contrast enhancement is to enable an image display device (such as a film plotter, a CRT display, or a computer line printer) to utilize its full dynamic range capability. Consider a 256-level plotter being used with 64-level Landsat MSS data. Unless some contrast stretching is done to the data, the resulting image will be too dark (since only the lowest fourth of the dynamic range of the plotter is being used). Another use for contrast enhancement is to bring out the information in a particular region in an image.

A standard method of stretching the contrast of a digital image is to apply a first-degree polynomial function to the image data. That is, let

$$C(i,j) = g\ I(i,j) + b$$

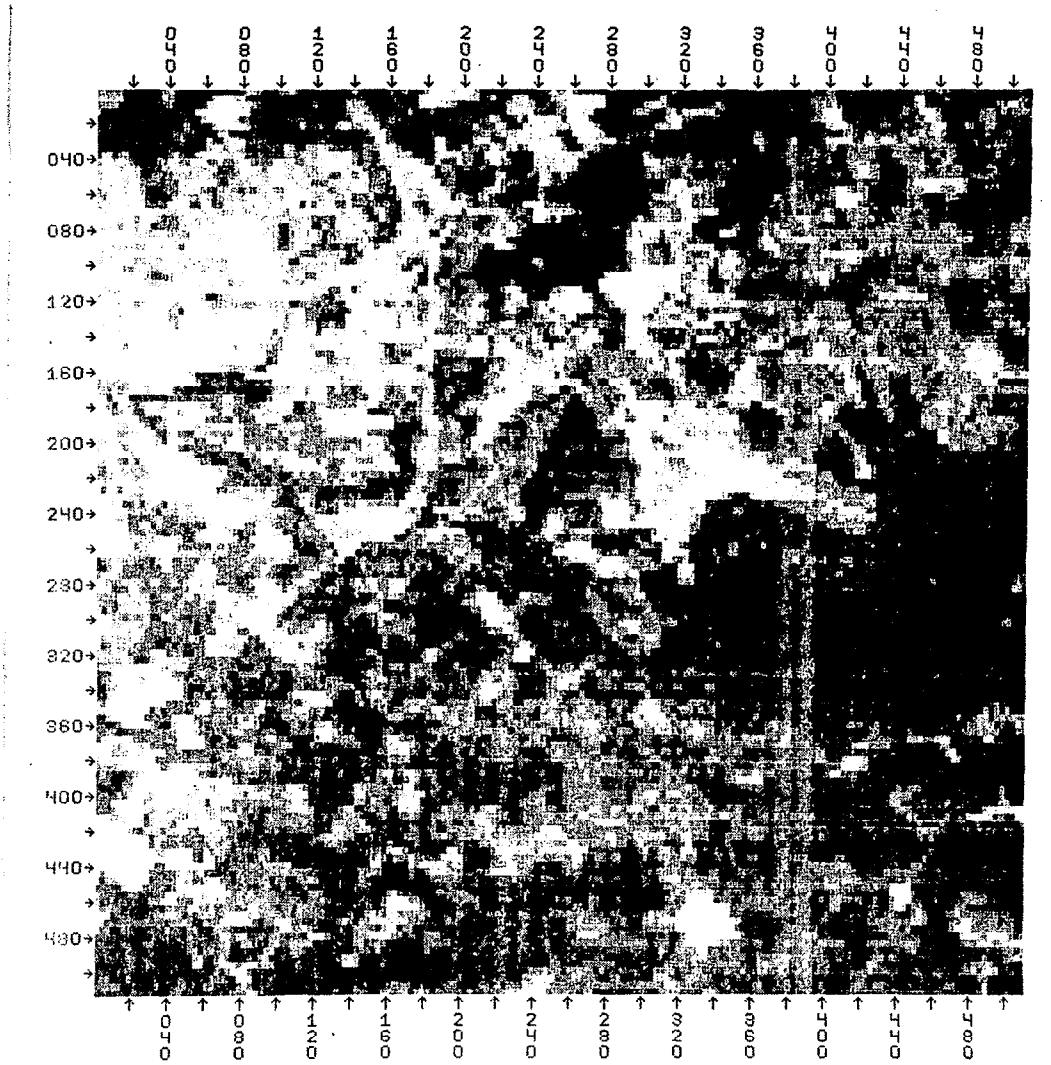


Figure 40a. Point duplication results as applied to Landsat Band 5 imagery from Salisbury, Maryland area

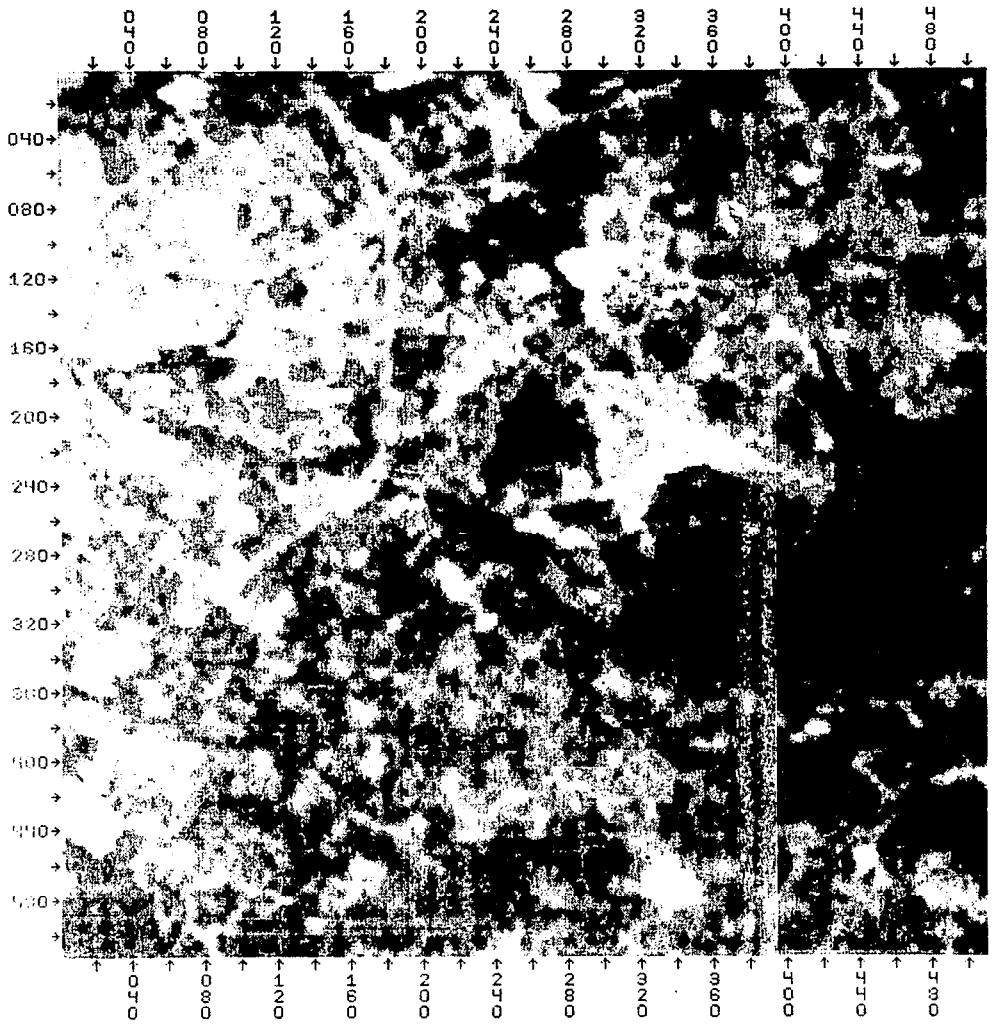


Figure 40b. Cubic interpolation results as applied to Landsat Band 5 imagery from Salisbury, Maryland area

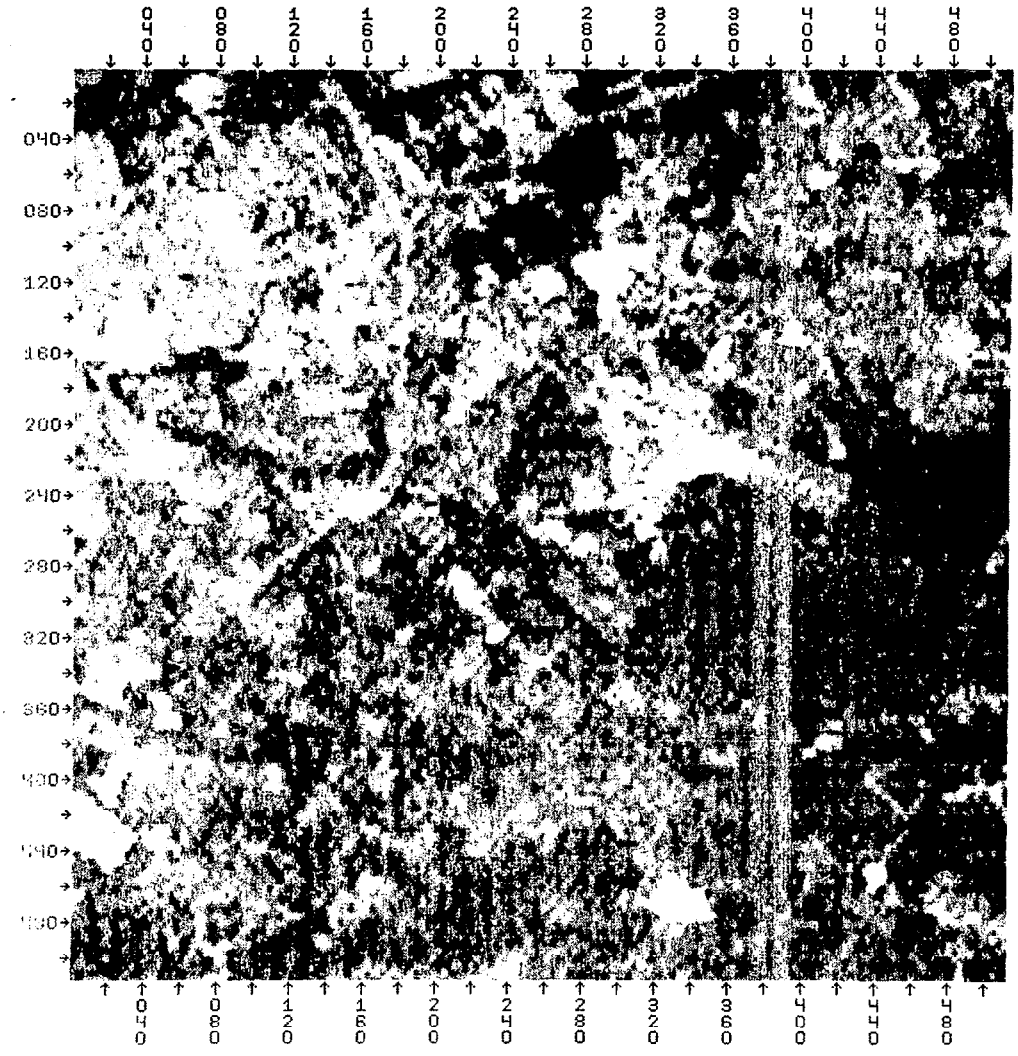


Figure 40c. Deconvolution algorithm results as applied to Landsat Band 5 imagery from Salisbury, Maryland area

where

$I(i,j)$ = intensity value of i^{th} sample in j^{th} line in the input da

g = gain coefficient.

b = bias coefficient.

$C(i,j)$ = intensity value of i^{th} sample in j^{th} line in the contrast-stretched data.

It is assumed that the values $C(i,j)$ have been clipped to fit the the display device.

The gain and bias coefficients can be determined through a very si technique:

$$g = S/_s$$

$$b = M - g\,m$$

where

m = mean of the subimage of input data that is of interest.

s = standard deviation of the subimage of input data.

M = desired mean of the subimage of output data.

S = desired standard deviation of the subimage of output data.

The values of the statistical parameters of the subimage of output mined empirically. However, reasonable values may be estimated by lett point of the effective dynamic range of the image display device and by sixth the length of the effective dynamic range of the device.

For example, consider a plotter that accepts pixel values in the r and that plots the same gray level for all values in the range 0,1,..., 56,57,...,255 assume that the response (gray level versus pixel value) linear. The effective dynamic range of this hypothetical plotter is 56 M=155 and S=33 should be reasonable values to obtain a good plot with t perience would enable these values to be refined.

Edge Enhancement

Edge enhancement is an artificial sharpening of the features of a is often used to accentuate boundaries of features or to bring out lin image. Pictures of edge enhanced images are generally more appealing

unenhanced images (although this is very subjective).

Digital edge enhancement can be performed by a discrete convolution function of the form

$$E(i,j) = \sum_{m=1}^{3} \sum_{n=1}^{3} I(i+m-2, j+n-2)F(m,n)$$

where

I(i,j) = intensity value of i^{th} sample in j^{th} line in the input data

F(*,*) = 3 x 3 matrix of filter weights

E(i,j) = intensity value of i^{th} sample in j^{th} line in the edge-enhanced data.

The particular algorithm depends only on the matrix of filter weights.

To perform edge enhancement the filter matrix F must define some kind of high pass filter. Four examples of Laplacian-type, high-pass filters are:

$$F_1 = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} \qquad F_2 = \begin{bmatrix} -1 & -2 & -1 \\ -2 & 13 & -2 \\ -1 & -2 & -1 \end{bmatrix}$$

$$F_3 = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix} \qquad F_4 = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 5 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$

Each of these examples has the property that the sum of the weights is one. This results in an image that is the sum of the original image plus a Laplacian-filtered image. Therefore, the resulting radiometry of an edge-enhanced image will be somewhat comparable (statistically) to that of the original image.

SYNTHETIC APERTURE RADAR IMAGE PROCESSING

Principles of SAR Operation

Introduction

A qualitative description of the basic concepts of high-resolution, side-looking radars will be presented here. Its intent is to provide a broad view of the principles of operation for those who have little or no familiarity with the subject. A more rigorous and detailed discussion of side-looking SAR is given in the Appendix.

Basic Side-Looking Radar Concept

Figure 41 illustrates the essential elements of a side-looking radar. Unlike aerial cameras which utilize light from the sun or flares, radars have their own antennas to provide a source of illumination. The transmitter generates short pulses of microwave energy which are directed by the antenna into a narrow, fan shaped beam at right angles to the flightpath. Thus a narrow strip of terrain is illuminated. Reflections of each pulse are sequentially produced by each range interval of the strip.

The portions of these reflections or echoes returning back to the aircraft are received by the original antenna and directed by the transmit/receive switch to the receiver. In the receiver a signal is generated with an amplitude dependent on the strength of the microwave energy received at any instant. That signal controls the brightness of a moving spot of light on a cathode ray tube. For each transmitted pulse there is one complete sweep of the spot across the tube. This sweep is recorded on a moving strip of photographic film to produce an image of the narrow strip of terrain. On the next transmitted pulse, the aircraft and film have moved forward a small distance and a slightly different strip of terrain is imaged. Ultimately, these side-by-side, sequentially recorded strips build up a complete image of the swath of terrain illuminated by the transmitter.

For simplicity, airborne film recording is assumed in the discussion and early side-looking radar systems utilized this approach. More current systems have also data linked the signals to a ground station for photographic film recording and for digital or analog direct viewing displays.

Planimetric Radar Geometry

It is apparent from Figure 41 that side-looking radar has a different image formation process in the along track and across track directions. In the along track direction features are recorded in sequence as the aircraft passes them. Thus, if the velocity of the photographic film is kept proportional to the aircraft velocity, the along track separation of features on the image will be scaled directly to their true along track separation.

However, in the across track, or range, direction there are distortions. In Figure

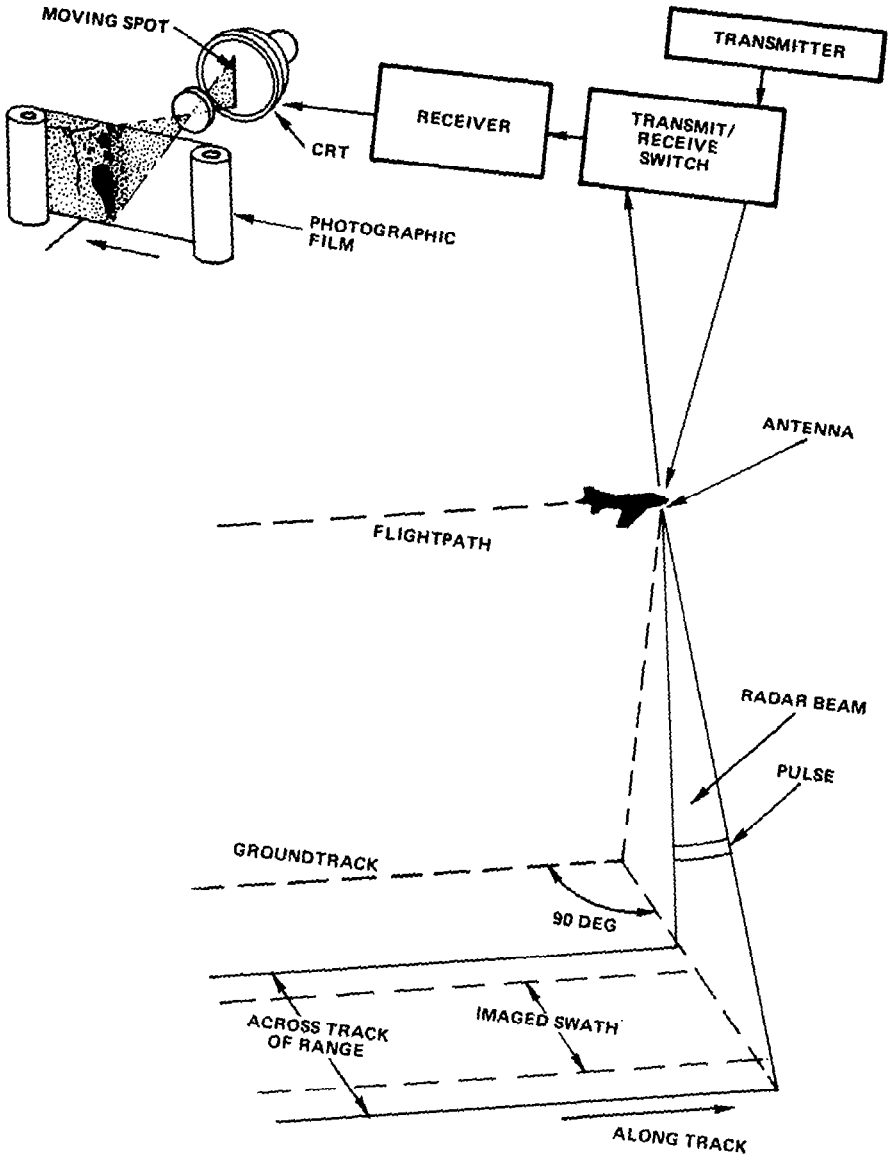


Figure 41. Side-looking radar concept

42a the bottom row of rectangles represents fields of equal size aligned in the across track direction. Since radar is a ranging device, it measures the difference in time between when the reflections from points 1 and 2 are received. This corresponds to their slant range interval, r, rather than their true ground separation, g. Thus the across track dimension of this field is significantly compressed on the imagery. On the other hand, the slant range and ground range intervals between points 3 and 4 are more nearly equal so the degree of compression is much less. For a fixed radar position the compression is greater at the nearer ranges. However, it is the cosine of the angle between the incident beam and the horizontal which determines the degree of compression rather than the range itself. Long and short ranges with the same incident angle produce the same scale compression. Notice that the radar results are in general the opposite of that produced by cameras. Vertical photographs have almost normal geometry while those taken to the side show oblique scale compression.

The across track distortions just discussed assume a constant velocity for the moving CRT spot in Figure 41. This produces a "slant range" sweep, where equal range intervals are displayed on the image with the same separation. If the absolute terrain clearance of the radar is known, it is possible to generate a nonlinear sweep velocity for the moving spot which removes the across track compression. Such a ground range sweep provides accurately scaled separation for features lying across track so long as no distortions occur due to height variations.

Range Resolution

As discussed previously, the short radar pulse is transmitted through the whole vertical radar beam. Different parts of this pulse are reflected by different objects. In Figure 42b, part U of the pulse was first reflected by point Z. Later, part V was reflected by point X. When the pulse reaches Y, there will be a reflection from that point. For radar to distinguish between two features aligned in the range direction, it is necessary that all parts of their reflected signals be received at different times. Otherwise they will appear to be one large object. For example, in Figure 42c, the reflected pulses from points X and Y do not overlap. These reflected pulses, or radar returns, will be recorded separately on the image so that points X and Y will be resolved. If the pulses were longer, they would overlap and be recorded as one large return.

Note that the length of the pulse is determined by the radar transmitter and is not affected by the distance of the features from the aircraft. In Figure 42c, the relationship of the two return pulses would be exactly the same with the aircraft at position A as it is at B. Thus the aircraft can move twice as far away and twice as high without changing the resolution in the slant range direction. However, the true ground resolution does depend on the depression angle between the horizontal and the line of sight to a given feature. In Figure 42d, the pairs of objects at X and Y are the same distance apart on the ground. Both are in the radar beam but those at Y

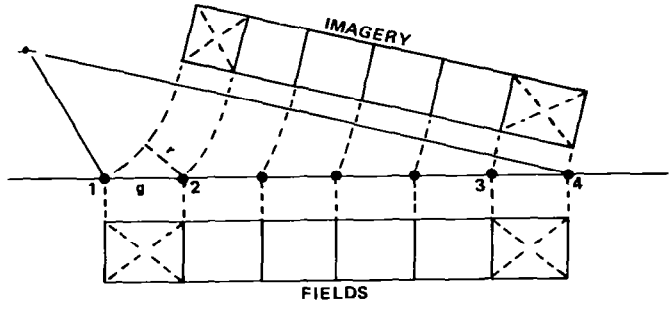


Figure 42a. Slant-range geometry

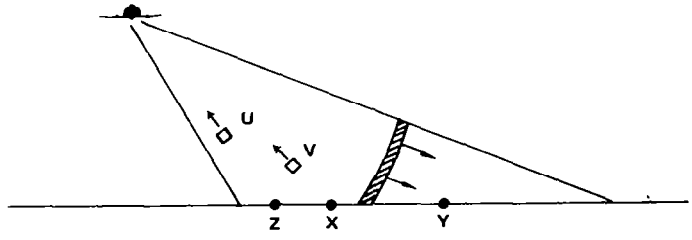


Figure 42b. Reflected pulse

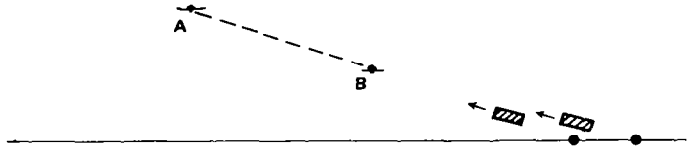


Figure 42c. Pulse length

Figure 42d. Range resolution

have a shallow depression angle, while those at X have a steep depression angle. It can be seen that the two pulses from Y do not overlap and will be recorded as separate spots on the image. The pulses from X overlap in range so they return to the antenna as one large pulse and will appear as a single strong return on the image. Thus, the radar resolution is poorest at steep depression angles when features are near the ground track and best at oblique angles. Essentially, the opposite is true for a camera. This discussion also illustrates why pulsed radars can not make images of terrain directly below the radar.

Pulse Compression

It is apparent that radar pulses should be as short as possible to obtain the best resolution; however, it is also necessary for them to transmit enough energy to detect the reflected signals. This means that if the pulse is shortened, its amplitude must be increased accordingly to keep the same total energy in the pulse. But the equipment required to transmit a very short, high-energy pulse is difficult to design and build. To overcome this difficulty, pulse compression techniques have been developed.

The first such approach was the electronic chirp technique. Instead of a short, constant frequency pulse, a long pulse with linearly varying frequency was used. The returns reflected back from this linear FM pulse were fed through a dechirp network for processing. This network acted as a matched filter, in effect delaying the different frequency portions of the pulse by amounts proportional to their frequency. Thus the long chirp pulse was compressed into a short, high amplitude pulse providing the desired improved resolutions.

In subsequent developments optical techniques were found to be more effective for storing and compressing the linear FM pulse. The frequency variations as well as the amplitude of the reflected pulse were displayed by the CRT spot (Figure 41) and recorded on film. When illuminated with coherent light, the linear FM film density variations focus to a point in the range direction to provide the compressed pulse.

Still more recently, digital techniques have been used for compressing the linear FM pulse.

Real Versus Synthetic Antenna Systems

For a conventional side-looking radar, the along track resolution is determined by the beamwidth of the real antenna as shown in Figure 43 . The angular width of the beam is constant so its along track dimension increases directly with range. Two objects at the same range can be resolved only if their along track separation is larger than the antenna beamwidth. Both objects at (b) are in the radar beam simultaneously, their reflections are received at the same time and they will appear on the image as a single large return. However, the objects at (a) with the same separation are never in the beam at the same time so they will be resolved on the image.

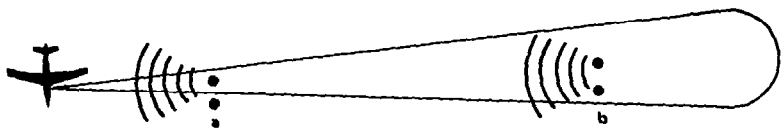


Figure 43. Real antenna resolution

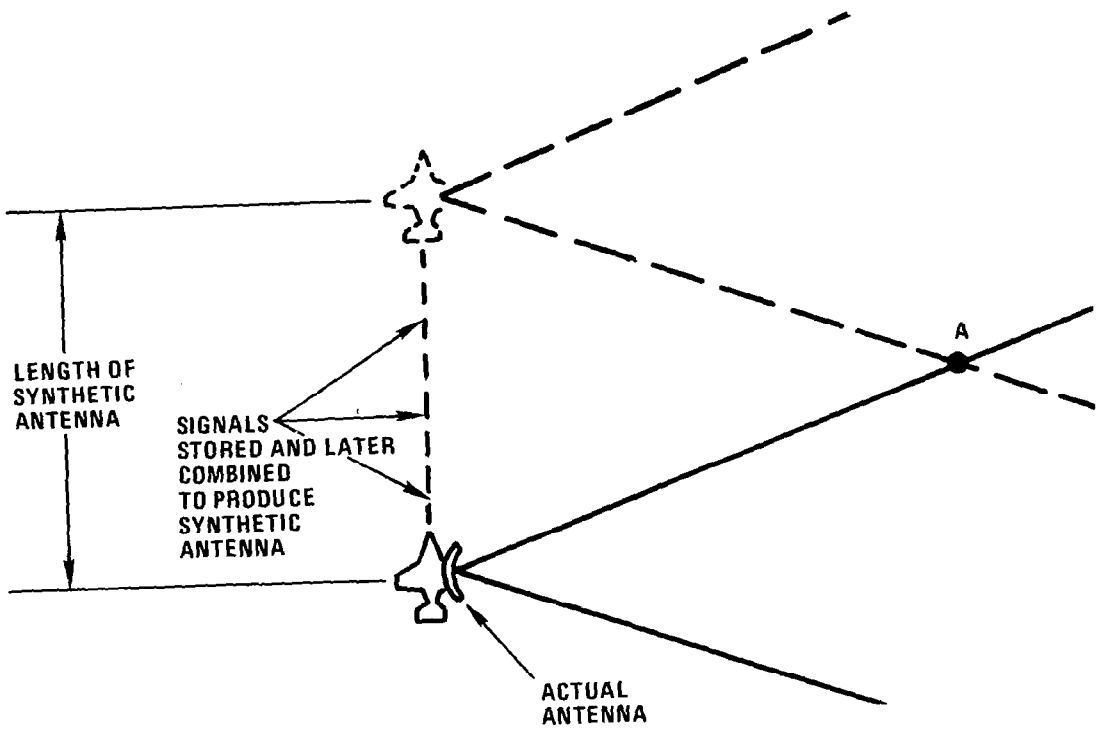


Figure 44. Synthetic antenna resolution.

The beamwidth of a conventional radar is inversely proportional to the length of the antenna. If the same resolution is desired at twice as great a range, an antenna twice as long is required. Even with the longest practical physical antenna, conventional radars are primarily limited to relatively short range applications. To overcome these limitations synthetic antenna radar systems have been developed.

The synthetic antenna, or aperture, is produced by having an aircraft with a relatively small actual antenna transmit and receive pulses at short regular intervals along the flightpath. When these individual signals are stored and then added, an antenna of long effective length is synthesized in space. In Figure 44, as the airplane flies past a feature at point A on the terrain below, the features enter the antenna's beam, moves through the beam and finally leaves it. During the entire time it is in the beam it reflects the series of microwave pulses. Now, the greater the range of the feature, the longer it remains in the beam of the antenna. In other words, the synthetic antenna is longer for more distant features and shorter for closer features. In fact, the effective length of the synthetic antenna for a feature is directly proportional to the range to the feature. Since along track resolution is proportional to the length of the antenna but inversely proportional to the range, the two effects compensate for each other, and the resolution remains the same at all ranges. Synthetic antenna radar thus makes it possible to obtain high-resolution images of terrain many miles away.

There is an important difference between the actual antenna and the synthetic antenna. For the actual antenna, only a single pulse at a time is transmitted, received, and displayed as a line on the image. Since all the signals received by the different elements of the antenna resulted from a single transmitted pulse, there must be a coherent phase relationship between them. For the synthetic antenna, however, each object produces a large number of separate return pulses. To provide a true analogy to a physical antenna of the same length each of the separate pulses must be coherent, in effect segments of the same continuous signal. The series of pulses from any given point on the terrain comprise the elements of the signal which must be stored and then coherently summed.

As a point moves through the radar beam its velocity relative to the antenna varies slightly producing a doppler frequency change in the series of pulses. By phase modulating the range gated pulses from the point against the same coherent reference used to generate the original transmitted signals, the continuous change in doppler frequency can be determined. This signal, commonly called the doppler phase history, is essentially a linear FM signal in the along track direction. It is closely analogous to the previously discussed chirp pulse being generated in the orthogonal range direction, except that it is changing at a much slower rate.

By appropriate sampling and processing in the radar's electronic units, and correct synchronization with the CRT sweep, the doppler phase history for each point can be

recorded in the along track direction on the film in Figure 41. In combination with the chirp pulse in the range direction, a Fresnel zone plate is produced for each point on the terrain (see Appendix). As with the chirp pulse alone, this two dimensional dispersed signal can be optically compressed using coherent light to produce an image film with fine resolution in both the along and across track directions.

More recent techniques have used digital recordings and matched filter electronic processing to eliminate the need for film recordings. Thus four basic synthetic antenna radar data sources are generally available: (1) Doppler phase history or radar signal data film, (2) Final image film, (3) Digital phase history data tape, and (4) Digital image tape for direct input to displays. The following subsection of this report discusses the practicality of converting one or the other of these data sources into an alternate form as required for particular applications.

Forming A Digital SAR Image

Digital recordings of radar imagery have the capability for much greater dynamic range than obtainable with optical correlation onto film from the same basic radar data. Because the radar system has a dynamic range typically of 50-60 db and film has a range only of approximately 30 db (that is a density range of 3), it should be possible to produce digital imagery with a quality that is superior, or at worst equal, to that obtained from optical correlations. To achieve this result, it is necessary to determine the optimum transfer function between the radar signal levels and the digital output brightness levels. This function has frequently not been adequately understood; resulting in the production of imagery of less than optimum quality.

Amplitude Coding Considerations

The optimum digital transfer function is one which produces equal intervals in log intensity, I, of brightness for equal intervals in log radar power, P_R, with a slope suitable to cover the range of radar signals of interest. Much of the available equipment which has been used to produce digital radar images does not provide this capability but instead gives equal steps in intensity rather than log intensity. Although some compensation can be made through the application of non-linear transfer functions to the output, the results will usually be less than optimum.

The Response of the Eye as a Receptor - For most applications, radar imagery, whether recorded digitally or optically is produced for visual evaluation. For this reason not only the important intensity information on the final output image must have a dynamic range tailored to the limits of perception of the eye, but also significant changes in received radar power must show up as significant changes in image intensity. The eye has a near logarithmic response to changes in intensity; that is, the eye perceives equal percent changes in intensity or brightness as equal changes. For example, a brightness change from 2 to 4 (arbitrary units) is perceived by the eye as the same

change as 4 to 8 and 8 to 16. Thus a brightness scale with equal logarithmic increments would be perceived by the eye as equal changes in brightness. This can easily be verified by observing two step wedges, one with equal intensity (I) steps and the other with equal log I (density) steps. Because the eye has a near logarithmic response to changes in intensity, the latter steps appear to have the same brightness change from step to step. However, the linear intensity steps appear to have a nonlinear brightness variation, with large changes between the darker steps and small changes between the brighter steps. The response of the eye has frequently not been given sufficient consideration by persons developing computer programs for digital recordings and displays. Because significant radar return changes are proportional to log P_R, a transfer function that maps log P_R into log I is appropriate.

Using Transfer Functions to Match Output to the Response of the Eye - The problem of finding the optimum transfer function was recognized when it was noticed that digitally produced SAR images generally were of inferior quality to optically produced images. It was found that the mechanisms of optical processing onto film automatically provided the ideal transfer function.

An optically correlated synthetic aperture radar image produced at the image plane of the correlator is one with a linear relationship between image intensity and reflected radar power. When this image is reproduced on film, the photographic process produces a density on the film proportional to the log of the exposure. Figure 45 shows a typical density (D) versus log exposure (log E) plot for a radar negative film transparency. Since the exposure is proportional to the image intensity, it is also proportional to the received radar power. The slope of the D/log E curve and the exposure are chosen so that the majority of the returns of interest fall on the essentially linear part of the curve. When a positive transparency is made from this negative, the density of the positive is inversely proportional to the density of the negative and thus to the log of the radar power. When the positive transparency is illuminated for viewing, the log of the intensity seen by the eye (log I) is inversely proportional to the density of the positive or directly proportional to the log of radar power (log P_R). Thus, the perceived relationships of the positive film transparency are those shown in Figure 46 . Since the relative image brightness as perceived by a viewer is proportional to log I, it can be considered that a linear transfer function converts the log of the received radar signal levels to observed brightness levels.

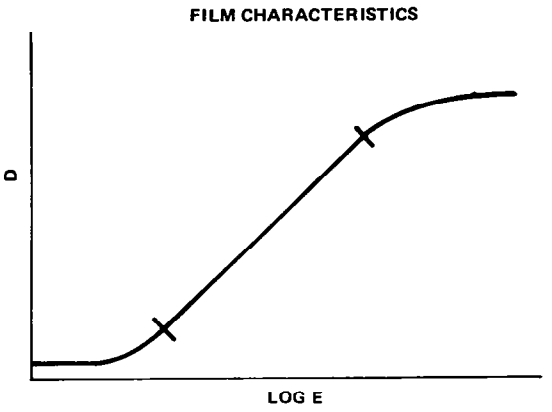


Figure 45. Negative film transparency characteristics

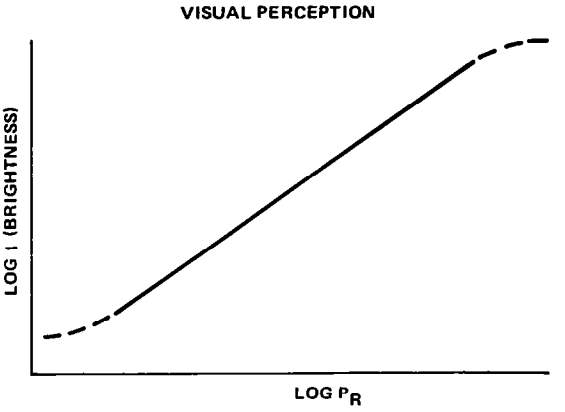


Figure 46. Perceived relationships for positive film transparency

To achieve similar results with a digital image, the perceived brightness or log of the intensity should be linearly proportional to the log of the radar power. The significance of maintaining a linear relationship between log I and log P_R can be illustrated with Figures 47 and 48. On both, plots of $I = P_R$ and $I = P_R^{1/2}$ are shown. In general, this relationship can be given either in the form $I = R_R^{\gamma}$ or the equivalent log form, $\log I = \gamma \log R_R$. The exponential γ simply determines the slope or contrast of the perceived display function. Above the maximum brightness, the signals are all clipped and displayed at the same intensity, since the film is saturate.

In general, the $I = P_R$ function produces an image with too much contrast; many of the weak returns are lost, most of the stronger returns are saturated, and there are very few gray levels between these extremes. The $I = P_R^{1/2}$ function produces many more gray levels and an image with better interpretability. From Figure 47, some observers in the past have mistakenly concluded that the square root function produces better results because the contrast decreases as the return signal level increases instead of the contrast being constant as for the $I = P_R$ function. From the log-log plot of Figure 48, the true perceived relationships are much more apparent. The maximum contrast is the same for both functions, but for $I = P_R^{1/2}$, the slope is one-half as great, providing a linear visual

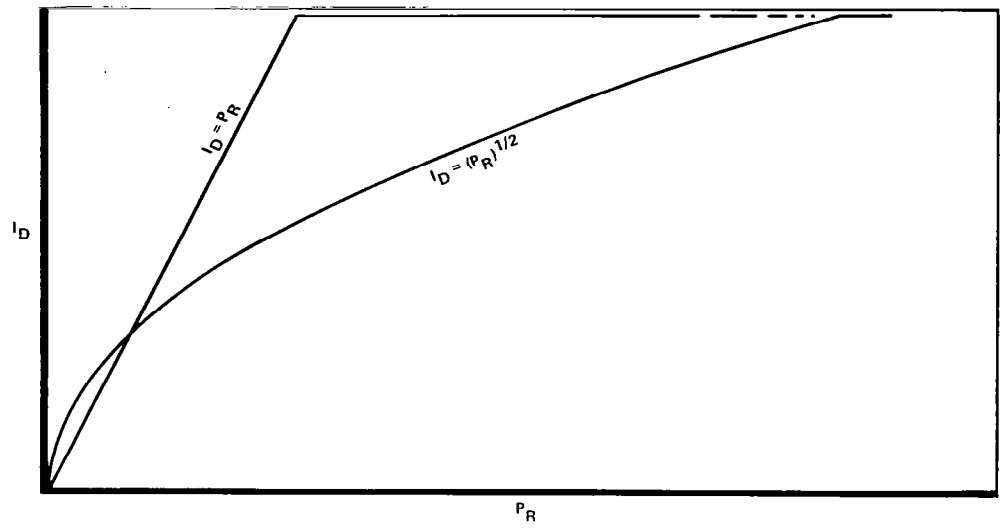


Figure 47. Linear plot of transfer functions

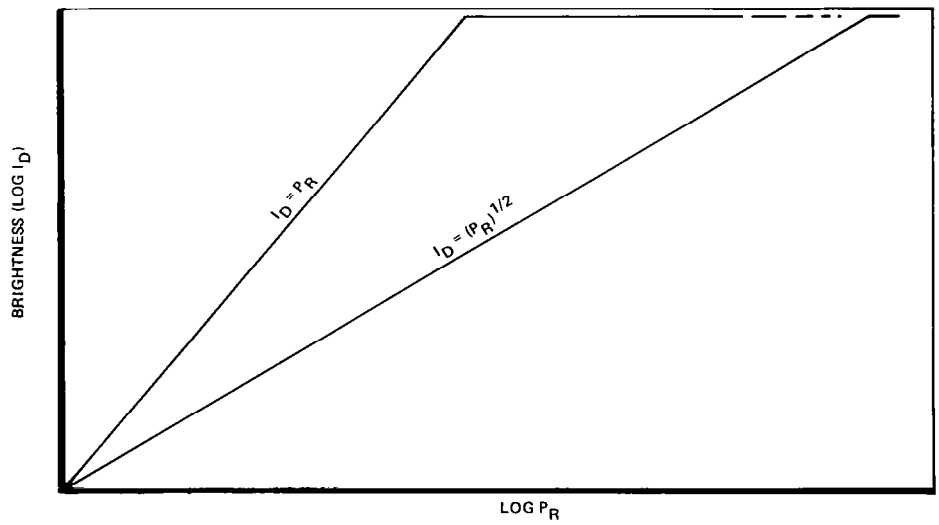


Figure 48. Log plot of transfer functions

response over a greater range of input signal levels, as long as a film image is made with log P_R mapping into log I. Of course, these results do not necessarily mean the square root function is optimum. Either a greater or smaller slope might be better. In general, the objective is to pick a slope which goes from the minimum signals of interest (the average of the shadow areas) to the maximum signals which can be clipped without serious image degradation. For the majority of radar imagery, this interval has been found to be no more than 50 db. To fully use the available dynamic range, the minimum brightness should be assigned to the minimum return signal of interest and the maximum brightness to the maximum signal of interest. The digital image, however, is often not transferred to film but rather is printed out using type symbols on computer paper. In this case as well as images printed by microdensitometer onto film, the brightness of the image is proportional to P_R or $P_R^{\frac{1}{2}}$ and not to their logarithms. This results in enhancing noise at low radar power returns and in compressed signal intensity for high returns. For a digital image the function will not be a continuous function but a series of discrete steps as indicated in Figure 49. To provide the closest analog to the optical correlation case, these steps should be small enough so their difference in brightness is less than the minimum discernible brightness level of the eye. In practice, this ideal is sometimes not achievable because of equipment limitations, but reasonably close approximations are often possible.

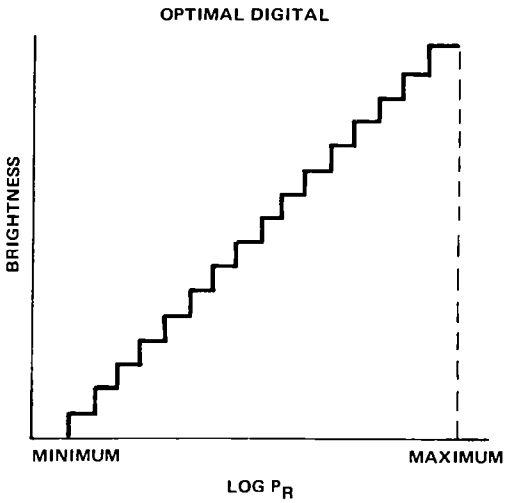


Figure 49. Equal log interval digital display function

Digitizing a Film Image Product

A digital SAR image may be obtained by digitizing a film image product with a microdensitometer or other digitizing system that can scan the image and output a digital value for each pixel or picture element.

If one has access to SAR imagery that is normally provided as a film transparency or print, the use of a digitizing scanner is recommended. The optimum scanning of the SAR image requires the consideration of image scale, scanning aperture, dynamic range of information on film, and digital values to be output. Given an image scale, the scanning aperture will determine the size of the picture element or, in other words, the resolution of the digital image. Since the radar image will often be registered to another image such as from Landsat, the scanning aperture size which determines the area covered by each pixel must be chosen so that the pixels are compatible, or a resampling algorithm must be available to bring the pixels into registration.

If the dynamic range that can be recorded on film is not appropriate for the experimenter, another approach must be used to obtain the digital SAR image. The alternatives when the SAR doppler phase history recorded on film is available are (1) to digitize the phase history and do a digital correlation in a computer or (2) digitize and record on magnetic tape the output of the optical correlator at the image plane.

In the case of AN/APQ-102A imagery, the first option requires a microdensitometer with a scanning aperture of 10 to 15 micrometers to fully realize the resolution of the radar. At this aperture large amounts of data are generated from the uncompressed phase histories and images of only a few square miles can easily take several hours to generate on a large general purpose computer.

The second option requires the optical correlation equipment, but rather than projecting the correlated image onto film, the image is projected onto an image dissector which gives a digital output of the image that can be stored on magnetic tape.

Distortions in SAR Imagery

Radiometric Distortion

Radiometric distortion is the result of deviations from the ideal radar transfer function and, if not compensated for, gives rise to errors in the measurement of surface reflectivities. The sensing geometry, e.g., introduces a distortion because of the increasing path length the microwave energy travels as the ground range increases. An antenna pattern that illuminates the swath in a manner that compensates for the range distance variation will eliminate this geometric radiometric distortion. Another technique to compensate for range variation is to increase the receiver gain as the returns from longer slant ranges are received with sensitivity time control.

Variations in transmitted power result in radiometric errors. Also, if the antenna is not correctly oriented for the desired swath (antenna roll error), the ground

illumination may not be correct, again producing radiometric distortion. The antenna roll angle and transmitter power can be monitored and, during processing, these measurements can be used to correct the distortion.

Vertical Antenna Pattern - It can be shown that if the vertical antenna pattern has a gain, G

$$G = K \csc^2\theta \cos^{\frac{1}{2}}\theta ;$$

where K is a constant and θ is the depression angle the range vector makes with the horizontal, then the received terrain backscatter will be independent of θ. The actual antenna patterns for the arrays used, however, will only be able to approximate the ideal gain function. The difference between the ideal and the measured patterns can be determined and used to make the needed small scale radiometric corrections in the imagery.

Sensitivity Time Control - The sensitivity time control (STC) adjusts the receiver gain to compensate for the antenna pattern variations as a function of slant range. The receiver gain becomes a function of the antenna pattern, altitude, and depression angle. The STC applies a predetermined gain variation for a given set of imaging parameters and any deviation between the actual and ideal STC functions can be used to correct the imagery during processing.

Geometric Distortions

Planimetric Distortions

Planimetric distortions are geometric errors in the imagery of level terrain. Elevation distortions are covered in the next subsection. Sources for planimetric distortion fall into two categories - (1) sensor geometry errors, and (2) radar and navigation equipment errors.

Sensor geometry errors are slant-range distortions inherent in the system. They occur because a targets position is determined by the time delay to receive the return signal along a slant range. This distortion can be corrected to give "ground-range" imagery by using nonlinear range positioning for imaging the return information.

Radar equipment errors occur in the across-track (range) and along-track (azimuth) dimensions. Across-track errors in a film recorder system are the result of deviations in the range recording sweep from the ideal sweep or of nonlinearities in the recorder film drive. Along-track errors result from residual errors in the clutterlock and motion compensation systems and from errors in measuring the motion which is being compensated.

The clutterlock and motion compensation corrections are range dependent and must be computed for each range sample. A simpler compensation scheme divides the range into channels and applies a correction calculated for a single range of the channel to the entire channel. The difference between the applied correction and the actual correction needed for a given range element is the residual error. Errors in measuring the platform's actual three-dimensional motion result in the application of the wrong compensation

to the range samples by the radar.

The platform's navigation system supplies the radar with ground speed and attitude information. Ground-speed errors cause aspect ratio and absolute position errors. The antenna roll and pitch angles are often determined by the navigation system. A roll error can result in a radiometric error, while pitch errors misposition the actual antenna beam which gives a geometric error.

Sensor Geometry Distortions - Images in a ground-range presentation are easier to mosaic and are in a format familiar to geoscientists. Distance measurement is simplified by the constant range scale. However, range information recorded linearly with time is in a slant-range format. Surface features appearing in the near range are more compressed than in the far range resulting in geometric distortion. Figure 50 illustrates the geometry and the resulting distorted display of surface features. Targets A, B, and C have equal ground-range dimensions, yet in the slant-range display, A' appears shorter than B' and B' shorter than C'. From the geometrical relationships:

$$\frac{A}{A'} = \csc i_A = \frac{R}{(R^2 - H^2)^{\frac{1}{2}}} = \text{secant of the depression angle}$$

Thus a ground-range presentation can be obtained by recording the data with a hyperbolic range sweep in a film recorder. The shape of the sweep waveform is a function of slant range (R) and the radar terrain clearance (H). The ground-range, however, is correct only for flat terrain at the specified clearance.

Radar and Navigation Systems Errors -

1. Across-Track Errors - Range or across-track errors are usually attributable to nonlinearities in the sweep of the beam of the film recorder for optically recorded data. Errors in the ground-range sweep deflection waveform result in range positioning errors.

(A) Aircraft Position - Because all measurements are made with respect to the airborne platform, uncertainties in its position and velocity result in errors in the final image. Therefore, accurate measurement requires the use of high quality navigation equipment in the aircraft. The basic navigation system normally used is the inertial navigation system (INS). This can be supplemented by Doppler navigators, ground control, and/or radio navigation equipment such as SHORAN or HIRAN.

In addition to the fact that uncertainties in the aircraft horizontal position along the average flightpath determine the overall position in the final image, errors in altitude and variations of position along the average flightpath produce internal errors in the image.

(B) Altitude Uncertainty - The equation for image coordinate x as a function of time to achieve a ground range sweep is

$$x = pR_g = p(R^2 - h^2)^{\frac{1}{2}} = p(\frac{c^2t^2}{4} - h^2)^{\frac{1}{2}}$$

where

x = sweep position
c = velocity of propagation
t = time
h = altitude above the datum plane
p = image scale factor.

Differentiation and some manipulation of the above equation results in:

$$\frac{dR_g}{R_g} = \frac{dx}{x} = -\frac{h^2}{R_g^2}\frac{dh}{h} = -\tan^2\theta\,\frac{dh}{h} \quad ,$$

where θ is the depression angle from the flightpath.

Thus, errors in aircraft altitude produce errors in across-track position which are small at shallow angles but large at steep angles. Errors of one-half percent or so are not uncommon in altitude measurement so that a nonlinear scale distortion of a percent or more might be expected. Of course, rangemarks undergo the same distortion so that measurements with respect to them are not affected, except for a second-order effect in interpolation accuracy. Rangemarks are normally spaced equally in slant range, and this should be taken into account for accurate interpolation. Thus, altitude uncertainty need not effect accurate range measurement.

Elevation Displacement - The above equation also predicts the amount of elevation displacement to be expected from objects above or below the datum plane. As in photography, the differential displacement from nonlinear flightpaths can be used to measure the elevation of terrain features. The accuracy of such a measurement is limited by system resolution and the knowledge of flightpath position. Also, a visual model may be obtained by viewing two such images stereoscopically, but an apparent overall warping is observed because of the differences between radar and optical elevation displacement.

2. Along-Track Errors

(A) Scale Errors - Several possible error sources in the total system contributes to an error in the along-track scale. The principal error is in the knowledge of aircraft velocity. Other possible errors are in converting this velocity to recorder signal film drive velocity and in synchronizing image film velocity to signal film velocity in the optical correlator. Of course, if sufficient ground control or accurate

navigation data are available, this error can be removed in data reduction or in image correlation. When an INS is used to measure aircraft position, the errors are periodic (Schuler induced), so a few ground control points can be used to determine the amplitude and phase of this cycle, and residual errors of 0.1 to 0.2 percent rms have been achieved.

(B) Pointing Errors - Most SARs are instrumented to map along a perpendicular to the flightpath averaged over several synthetic apertures. Thus, if the navigation system makes long-term errors, such as the Schuler error, errors in both image straightness and orthogonality will occur. These are usually small compared to other errors discussed here, and can be improved with moderate use of ground control.

Probably the largest source of error in SAR is the accuracy of pointing the synthetic beam with respect to the nominally straight line flightpath.

The control is usually accomplished with a clutterlock which uses the radar signals to provide synthetic antenna pointing, because the radar can perform an accurate measurement of the Doppler frequency shift of the signals caused by relative motion between the

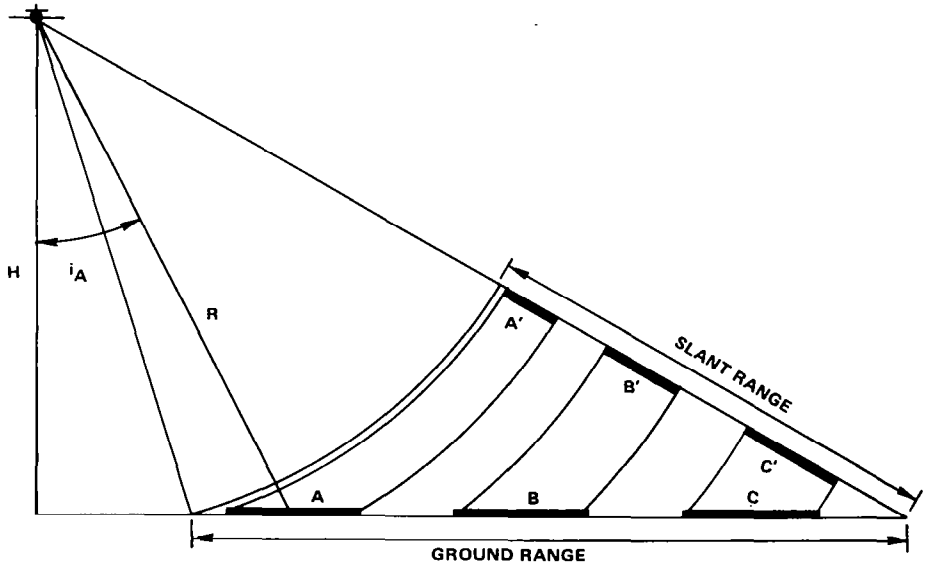


Figure 50. Slant range distortion

aircraft and terrain scatterers. As shown in Figure 51 the antenna beam will be in general misaligned by some angle, ψ. The Doppler shift in the signal from a scatterer in the center of the real antenna beam is $f_D = 2v/\lambda \sin \psi$. Electronic circuitry in the radar, referred to as a clutterlock system, effectively measures the energy in the signal having positive Doppler shift and balances this against energy having negative Doppler shift, providing a signal to either (1) drive the antenna so that the beam is aligned with the zero Doppler line (normal to the flightpath) and/or (2) offset an electronic reference to receive and process signals in the center of the physical beam, thus steering the synthetic antenna.

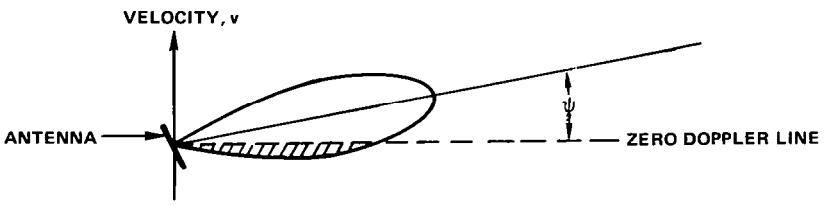


Figure 51. Clutterlock

Clutterlock accuracies in beamsteering are usually on the order of one-tenth the physical beam. For a one-degree beam, this amounts to about 0.1 degree.

This is an rms value averaged over a time period and may be exceeded for short periods. For example, the clutterlock errors can build up to be several times this value when land/water boundaries, or other large variations, cause half the antenna beam to contain large scatterers compared to those in the other half. This can be minimized by proper choice of system time constants.

When the flightpath is not horizontal momentarily, the synthetic beam will lie entirely ahead of or behind the normal to the average flightpath. This error is usually comparatively small because the (pitch) angular error is the same magnitude or smaller than the yaw error, but the lever arm (altitude, perhaps 10 kilometers) is

much smaller than the maximum lever arm for yaw errors; i.e., maximum range of 50 kilometers. Thus, the error would not be expected to be as high as 20 meters.

Another error source which is normally small but occasionally must be considered arises when a clutterlock error or synthetic beam correction causes the synthetic beam center to be on a Doppler cone rather than in the vertical xz plane. Because $f_d = 2v/\lambda \sin \psi$, the locus of constant f_d for $\psi \neq 0$ is a cone whose axis is the flightpath and whose intersection with the datum plane is a hyperbola. The geometry is shown in Figure 52.

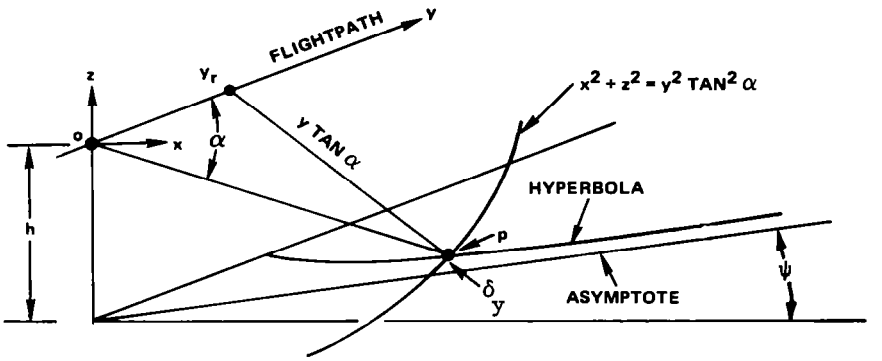


Figure 52. Doppler cone strategy

In the figure, line o-p is an element of the Doppler cone whose semiapex angle is $\alpha = \pi/2 - \psi$. Point p lies on a circle with center 0 at $y = y_r$ parallel to the x, z plane

$$x^2 + z^2 = y^2 \tan^2 \alpha ,$$

which intersects the datum plane $z = h$, so that

$$y = \frac{(x^2 + h^2)^{\frac{1}{2}}}{\tan \alpha}$$

is a hyperbola in the datum plane.

For large x, the hyperbola is asymptotic to the line $y = x/\tan \alpha$, so that the error between the position of imaged point p and the assumed position in the plane is

$$\delta y = (x^2 + h^2)^{\frac{1}{2}}/\tan \alpha - x/\tan \alpha \ .$$

For example, if h = x = 40,000 ft, and α = 90 deg - 0.1 deg = 89.9 deg, then δy = 28.9 ft.

It should also be noted that an elevated terrain target having the same range, R, and Doppler angle, α, as a point in the datum plane must also lie on the same circle and the elevation displacement is always toward the flightpath in a direction normal to it.

Elevation-Related Geometric Distortions

This section discusses distortions, or more accurately, displacements, and shadow formation of natural and cultural features that are either elevated or depressed.

Shadows - in utilizing its own illumination, radar systems produce "shadows" unique in character and construction that are completely independent of sun angle. Radar shadows are those areas of a display not illuminated by the incoming microwave energy because of obstruction by features such as hills or large buildings.

Because of its longer wavelength, radar energy undergoes no significant scattering by the atmosphere. Thus, unlike photos, where scattered light reveals many features in shadow areas, radar shadow areas are totally black. Figure 53 shows the effect of a hill on a radar image. Energy striking the front slope of the hill (PT) is reflected diffusely with some energy returning to the antenna. The back slope, TS, and the terrain from S to V receive no energy to reflect so the corresponding areas on the film remain unexposed.

Radar shadow length depends on the depression angle to the top of the object generating the shadow. The shallower the depression angle, θ, measured from the horizontal, the longer the shadow. Figure 54 shows the effect of shallow and steep depression angles on shadow length. For a shadow to be produced, the depression angle of the slope must be greater than the depression angle of the radar. If it is less, the terrain will be illuminated and no shadows will be produced. In figure 55, incoming energy illuminates the rear slope, TS, so some energy will be reflected back to the antenna to produce an image.

For the unique data collection parameters of SEASAT, radar shadows from natural features will be virtually nonexistent primarily because terrain slopes greater than 70° are extremely rare. Probably only such unique features as mesas, cliffs and canyons, like those in the western United States, the fjords of Scandanavia or the ice cliffs of the Arctic or Antarctic are big enough and steep enough to generate radar shadows.

Cultural features will also produce few radar shadows under SEASAT conditions. Because of the steep depression angles even the tallest structures will produce only very

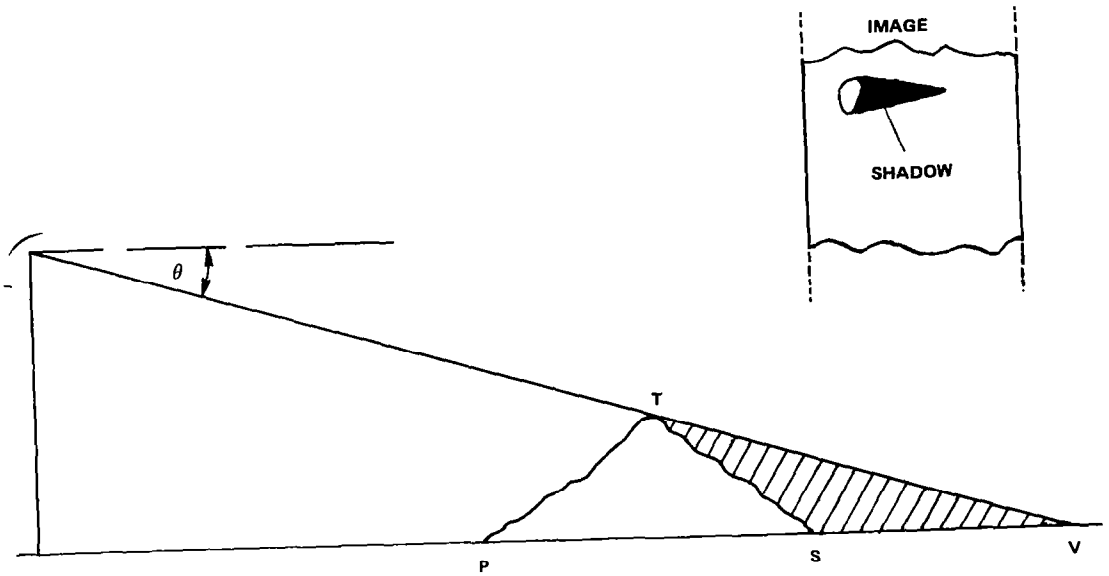


Figure 53. Formation of radar shadows

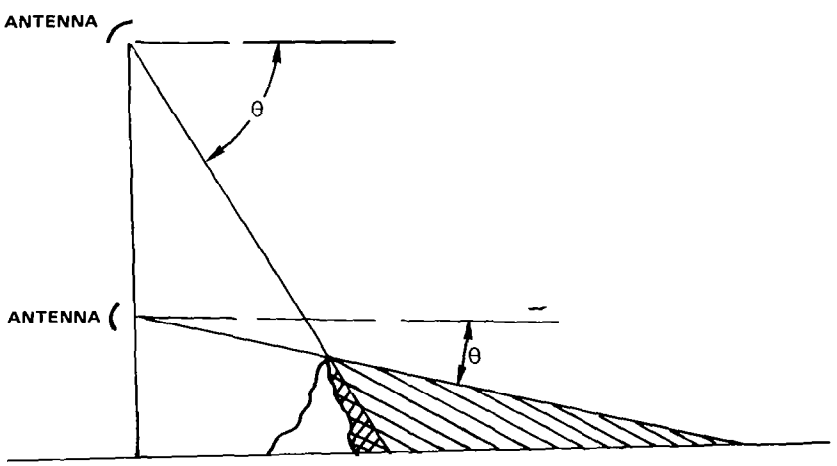


Figure 54. Radar shadows as a function of depression angle

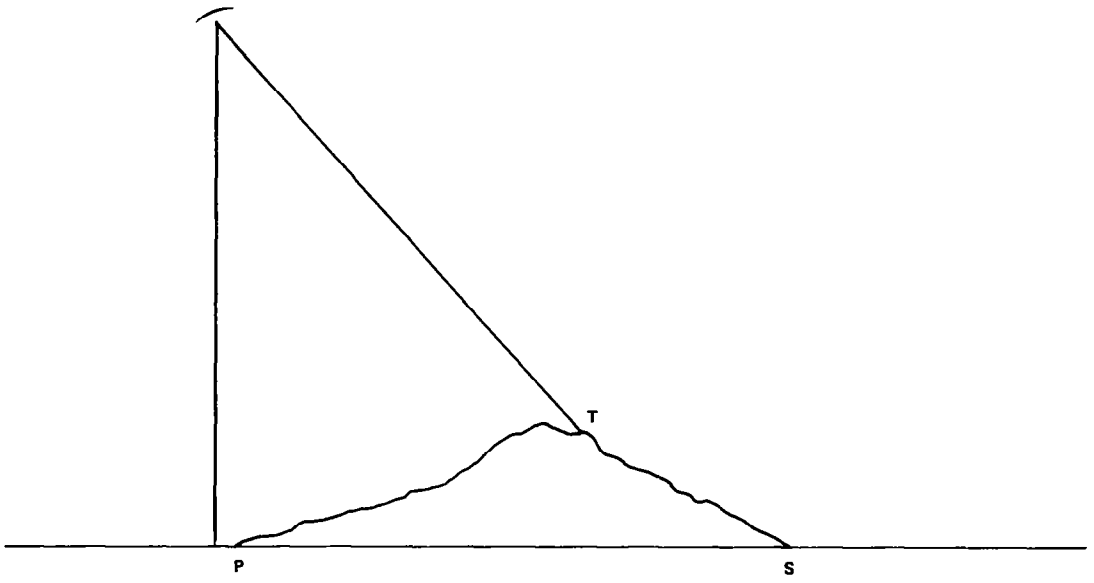


Figure 55. Effect of terrain slope on radar shadows

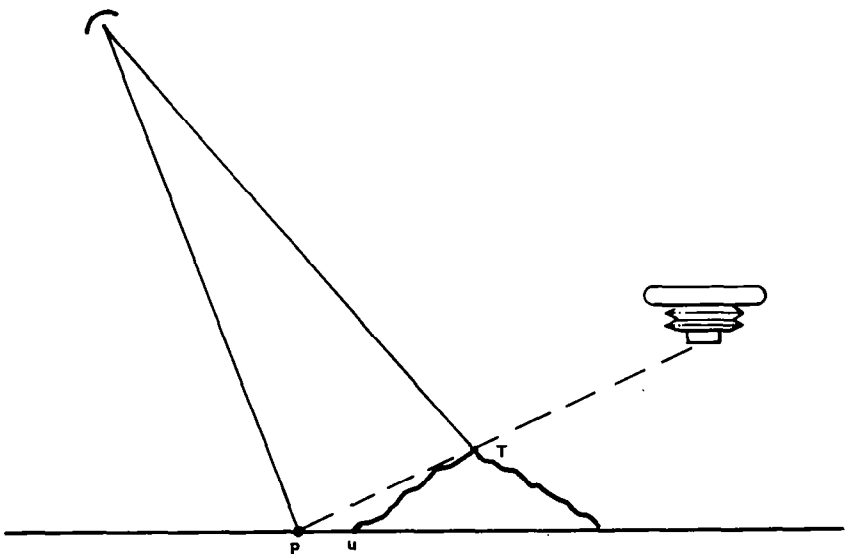


Figure 56. Radar and camera displacement of elevated features

short shadows. For example, a 100 meter structure will generate a shadow length of about 1.5 resolution elements. As with natural features, very few structures meet the requirements to produce appreciable shadows on SEASAT imagery.

Displacement and Compression of Elevated Features - Radar records and displays detected features as a function of the time required for transmitted energy to strike a target and return to the antenna. As a result, the top of a tall feature, being closer to the antenna, can be recorded and displayed at nearer ranges than its base.

This of course, differs from the aerial photographic condition. A camera records its images as a function of the angle between the optical axis and the target and so displaces elevated features away from the flightpath. Thus, the image displacement in radar imagery approximates that of an aerial photograph taken from the opposite side of the elevated feature, as shown in Figure 56. Points P and T are at the same range from the radar antenna and so will be recorded at the same time. Point U is farther away and will be recorded after P and T. For the camera condition, point T is displaced to P and U is not visible.

There is a fairly common misconception that converting from slant to ground range sweeps will in some way reduce or minimize elevation distortions or layover. It can be seen from Figure 56 that this is not true. Regardless of the type sweep used, returns from point T and P will be received simultaneously and recorded at an identical location. With a ground range sweep the horizontal distance $\overline{PU}$ will be in true ground range scale while with a slant range sweep it will be compressed. However, in either case, the relative location of T with respect to P and U does not change.

Obviously the depression angle and the slope of the terrain will determine the amount of displacement, or "lay-over", towards the antenna. The greater the depression angle the more a tall feature will be laid over. In both examples in Figure 57 the top of the feature, T, had been laid over to P. Points T and P, in each example, are at the same range and so will appear at the same point on an image. In Figure 57a, the shallow depression angle makes slant range and ground range to the target almost the same producing only a small amount of layover, PQ. In 57b the slant range is much longer than ground range and gives a greater layover.

In the ground range plane the layover, L, is related to the height, h, of a feature and the depression angle θ, by $L = h \tan \theta$. For a digital image which is to be registered to another image or map such a displacement can cause a problem. Vertical structures of one resolution element with depression angle of 70 degrees will have their tops displaced almost 3 resolution elements.

To be detected and measured by a photointerpreter (optically) the layover of a building or hill being illuminated must be four or five resolution elements. For the SEASAT system then, with its 70 degrees depression angle, this means the feature must be at least 34 meters high.

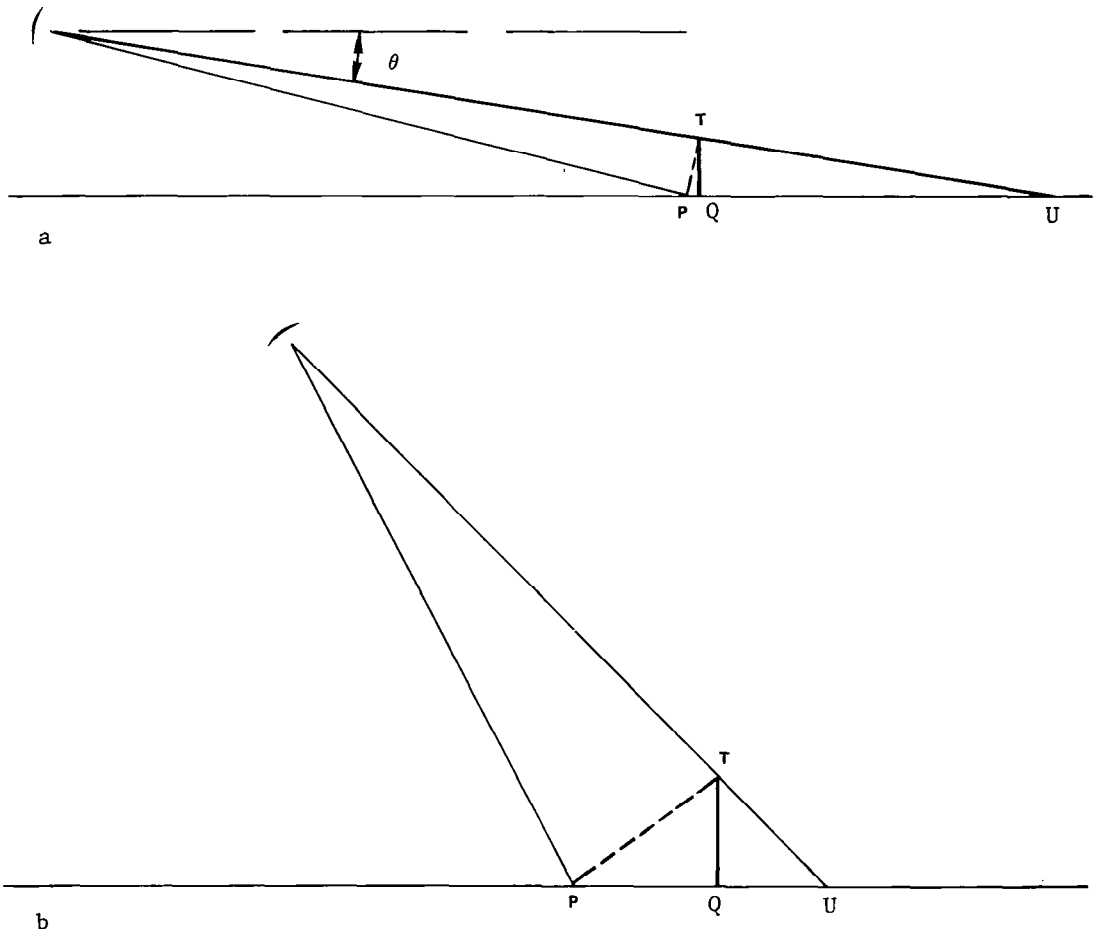


Figure 57. Effect of depression angle on range layover

Also in Figure 57b, assuming returns from the whole vertical surface, $\overline{TQ}$, the layover image in the ground range direction will be $\overline{PQ}$. This image length is recorded over the idealized shadow length $\overline{QU}$. The effective recorded shadow length $\overline{QU}$, for a structure of height h will be $\overline{QU}$ = h tan (90 - θ). With the 70 degree depression angle, a 100 meter structure will produce a shadow length of only 1.5 resolvable elements.

Depending on slope and depression angle, nonvertical features, such as hills and mountains, can be laid over or merely compressed in the across-track direction. In Figure 58, the angle γ made by slope $\overline{ab}$ is less than (90 - θ). All points on the front side of the slope are compressed much more than the back slope because points on $\overline{ab}$ are illuminated at nearly the same time or range from the antenna. In Figure 59, the angle γ is greater than (90 - θ). Point b is recorded first, followed by number 2, 1 and then a. The back slope $\overline{bc}$, does not undergo this same compression and reversal due to layover because the successive points, progressively farther on the ground from b are also progressively farther from the antenna.

In conclusion, shadowing and image displacement can cause interpretation and analysis problems, particularly in registering digital images. Vertical structures of one resolution element with a slant range depression angle of 70 degrees will have their top dis-displaced almost 3 resolution elements.

Correcting Errors in SAR Imagery

Radiometric Considerations for Aircraft Systems

Evaluation of Radiometric Variability - Aircraft SAR data can contain radiometric variability due to a number of causes. The primary variation is a function of slant range with the near image brighter than the far image. There is usually a systematic correction for this effect, however. Other errors are due to AGC effects, system noise, and film recorder variations both on board the aircraft and in the correlation process. The approach taken in this study was to analyse the resulting image data set for across track and along track variability rather than treat individual error sources separately.

This approach assumes that the radiometric errors are orthogonal in the along track and across track directions and thus can be analysed separately. The digital radar image is stored as an array of numerical values with the columns representing the across track dimension and the lines representing the along track dimension. All the data values in a particular column represent a particular ground range and all values in a line represent a particular along track distance.

Another assumption is made which is scene dependent and must be employed with care; that is that the scene is heterogeneous. This means that the scene classes are distributed uniformly over the area encompassed by the scene so that there are no large areas of one class. This is to say that if the scene contained no radiometric distortions

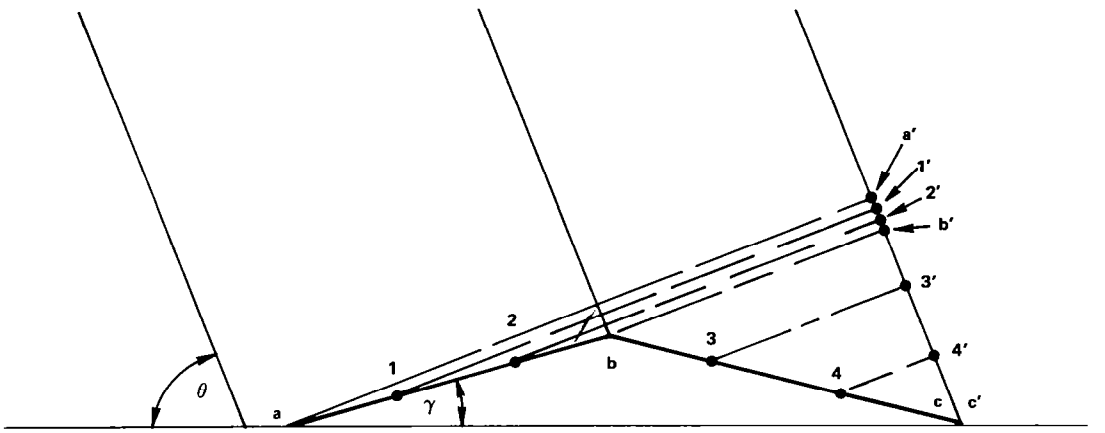


Figure 58. Displacement and compression for shallow terrain slopes

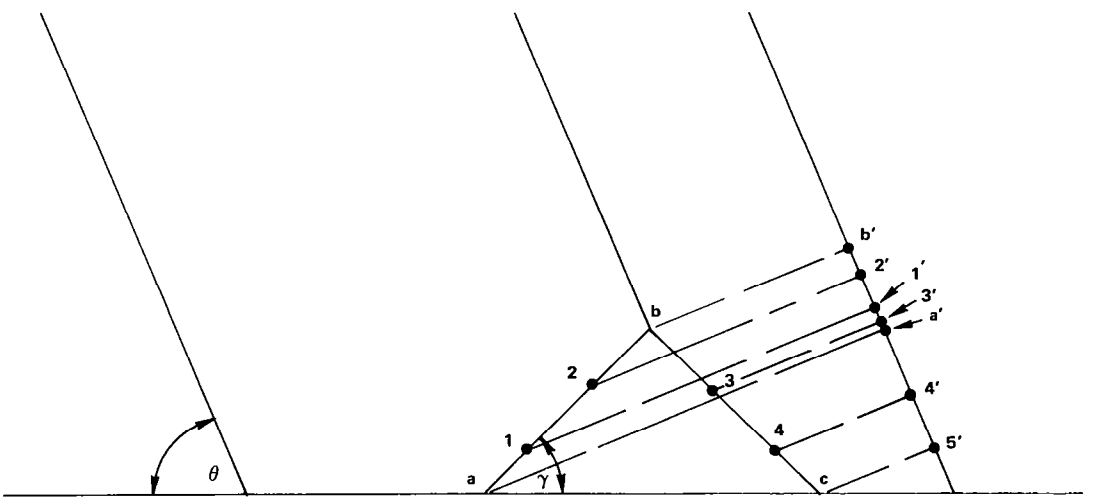


Figure 59. Displacement and compression for steep terrain slopes

that an average taken over one scene sub-area would be the same as that taken over any other sub-area. The statistical equivalent to this assumption is that the data itself is stationary and any nonstationarity in the output data is the error to be removed.

The radiometric analysis and correction procedure used computes orthogonal averages over the digital radar image and uses these averages to normalize the data so that the result is uniform or "stationary." This is done by computing an average for each column and for each line in the image as follows:

$$CM_i = \sum_{l=1}^{NLN} x_{ij} \qquad j = 1, NCOL$$

$$LM_i = \sum_{j=1}^{NCOL} x_{ij} \qquad i = 1, NLN$$

where: x_{ij} is the radar image brightness value at column j and line i.

NCOL is the number of columns in the image.

NLN is the number of lines in the image.

CM_j is the column mean for column j.

LM_i is the line mean for column i.

The column and line means are then used to normalize each column and line individually. A standard mean is chosen (SM) and each data value is corrected first across the track and then along track as follows:

$$\hat{Y}_{ij} = \frac{x_{ij}}{CM_j} \times SM \qquad j = 1,NCOL, \quad i = 1, NLN$$

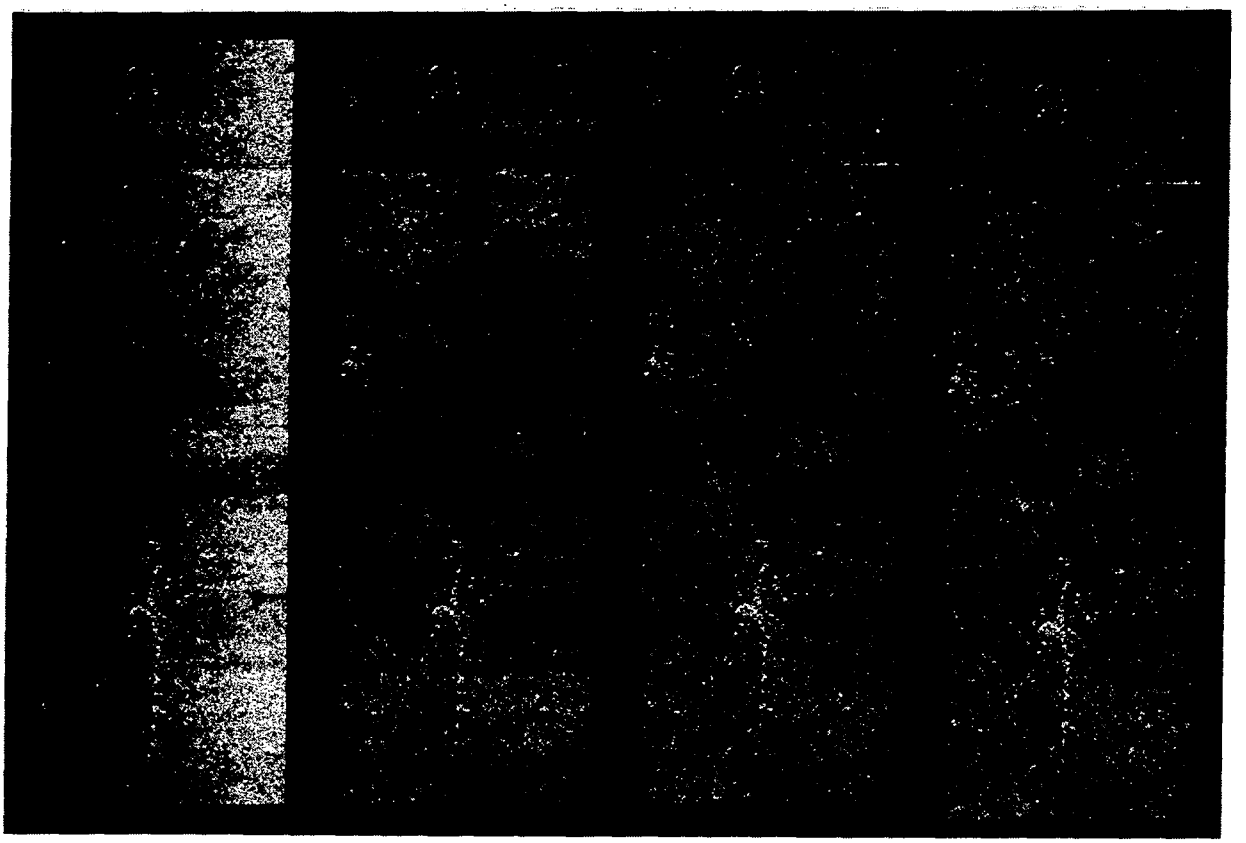
$$Y_{ij} = \frac{\hat{Y}_{ij}}{LM_i} \times SM \qquad i = 1, NLN, \quad j = 1, NCOL$$

where: $\hat{Y}_{ij}$ is the across track corrected radar image value for line i and column j.

Y_{ij} is the line and column corrected image value for line i and column j.

SM is a standard mean value.

A radar image which satisfies the assumptions corrected in this manner will have a uniform appearance with respect to radiometric errors. This approach was tested on the imagery from a SAR flight over Salisbury, Maryland. Figure 60 contains images made from the digital SAR data. The left figure labelled "original" is seen to contain severe left to right shading which distorts the image and makes it difficult to relate fields on one side of the scene to those on the other side. Other dark bands exist which run along



Original | Across Track Corrected | Along Track Corrected | Filtered Data

Figure 60. Radiometric correction example using SAR data from Salisbury, Maryland

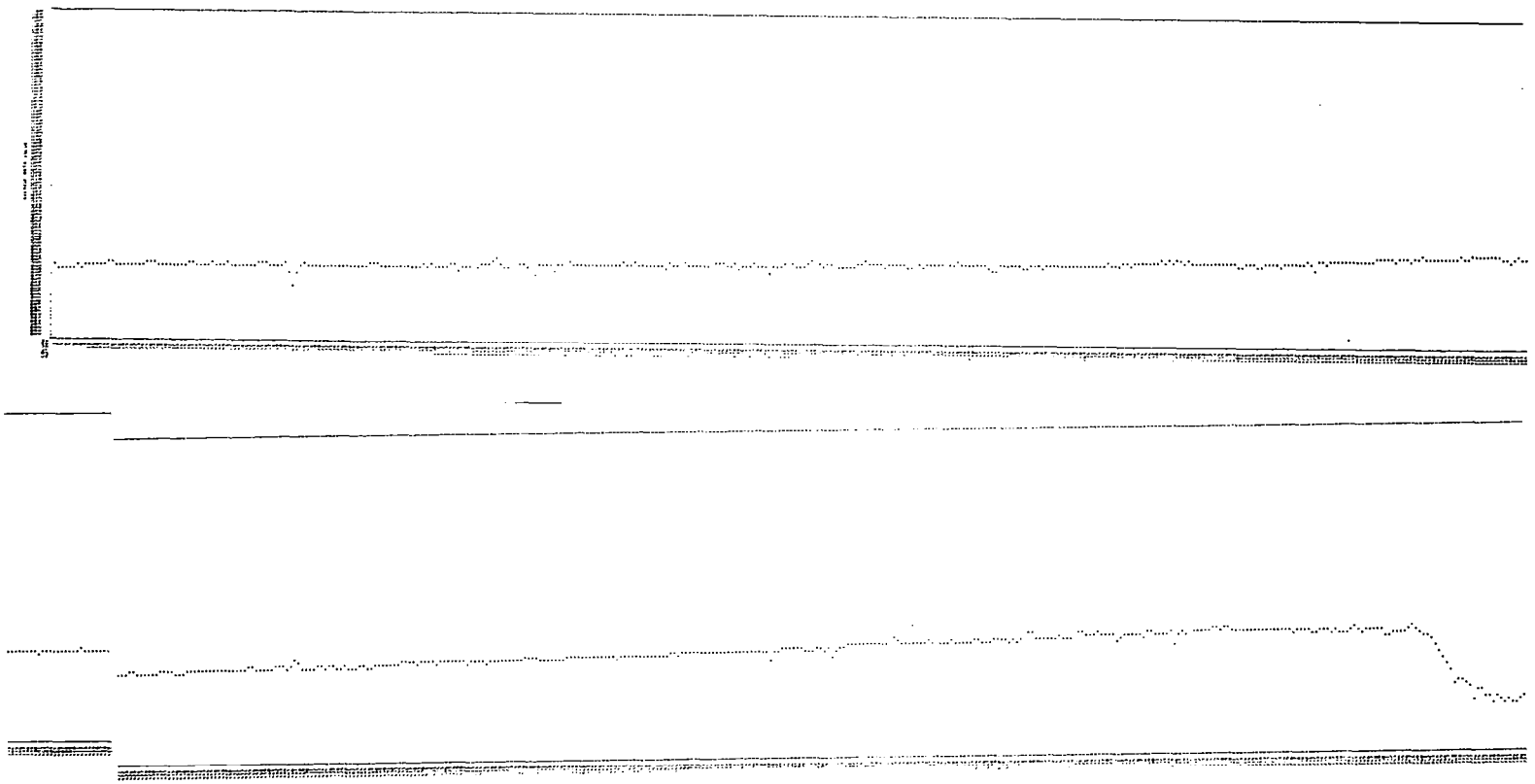


Figure 61. Plot of column means for Salisbury, Maryland SAR test data. The left side of the image is column one. Note the rise from left to right which corresponds to the increase in brightness in the original image in Figure 60.

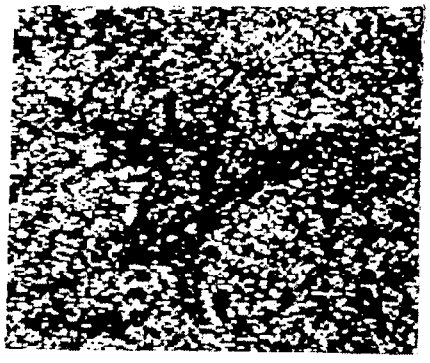
track. In the along track direction several dark bands are seen running across the image and a very dark narrow line exists near the top of the image running across the image. These distortions appear to be orthogonal across and along the image thus apparently satisfying the first assumption. The scene was also judged to be uniformly composed in that there were no large bodies of water or landforms in one area and not in other areas of the scene. The analysis and correction procedure was then applied to the data set.

Radiometric Correction using Line and Column Means - Figure 61 contains the column averages for the image. The line average graph is essentially flat and very long and was not reproduced here. The general brightness increase from left to right is seen in the column mean graph and the dark bands and very dark line errors can be seen in the line graph plot. The correction procedure for columns produced the second image from the left in Figure 60 and the along track correction applied to the across track corrected scene is presented in the third from left image. It can be seen that the across track shading and along track banding is satisfactorily removed. The correction could, of course, be done in one step but was done separately here to illustrate the effect of each correction. Note that some light shading exists above the right of the very dark error line near the top of the image and a dark area exists near the left end of the line. This is a correction error due to the non-orthogonality of the error; that is, the line is skewed in the image and the averages computed for the lines containing the error do not truly describe the line uniformly from the left to the right of the image. This is an example of the requirement for orthogonality of radiometric errors which must be met for this method.

The line and column averaging method of correction appears to be a feasible approach for correction of radiometric errors in SAR imagery at least from the aircraft system studied here. Since the method is context based rather than model based it is expected to work equally well on data from other SAR systems. Model based approaches for the AN/APQ-102A for example would have to correct for errors in the Sensitivity Time Control and Antenna Illumination pattern compensation. The line-column mean approach will allow users to correct for these and other errors by observing the data rather than fitting a set of models.

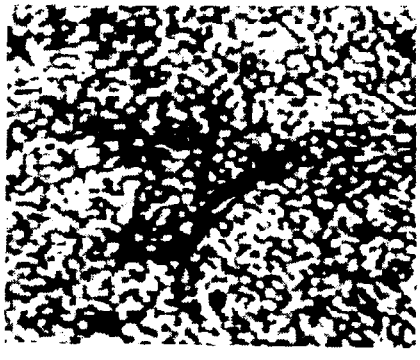
Two Dimensional Filtering For Noise Reduction - Observation of image reproductions of the Salisbury SAR data revealed a great deal of noisiness which made recognition of scene features difficult. It was decided that some form of image filtering may be desirable in the system if improvement in the signal-to-noise ratio would result. Experimentation was carried out on the Salisbury data to visually observe the effects of filtering.

A sequence of low pass frequency domain filters were tested using a two-dimensional FFT algorithm. Circular filter windows cutting off at frequencies as high as 78% of the Nyquist frequency and as low as 16% were applied to Fourier transforms of subsets of the scene. Figure 62 contains image reproductions of the original and filtered

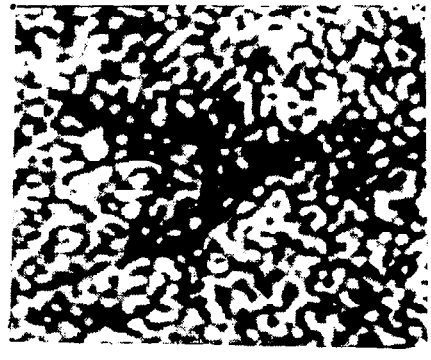


Original Image

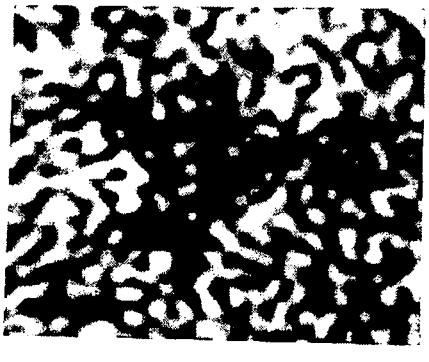
Cutoff Ratio F_c = .78



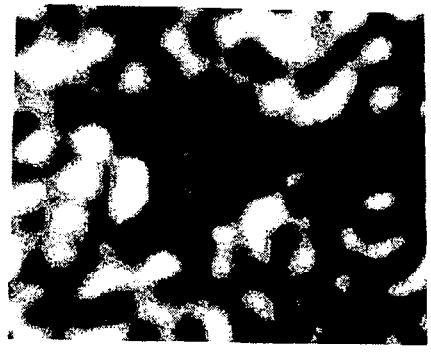
F_c = .63



F_c = .47



F_c = .31



F_c = .16

Figure 62. Examples of low pass filtered SAR imagery over Salisbury, Maryland airport

data. The original image is in the upper left position. The subimage is centered on the Salisbury, Maryland airport. The cutoff frequency fractions refer the cutoff frequency to the maximum, or Nyquist, frequency assumed to exist in the image. The subimages used here consisted of a 128 by 128 point block thus the maximum frequency is 2/N or 1/64. For example the 16% or .16 filter passed only the first ten frequencies and set all higher frequencies to zero.

Successive images in Figure 62 represent lower cutoff frequencies and it can be observed that the speckled noise effects become smoother and at the same time the runway in the image blurs out to larger dimensions, both expected results. At the .16 cutoff level the image is seriously blurred and clearly not enough high frequencies are present. On the other hand the .78 cutoff ratio produces very little change in the nature of the image. Visual evaluation suggested that a cutoff somewhere in the .47 to .78 range would smooth the image noise without unacceptably degrading the image.

The FFT filtering approach is cumbersome for larger areas and core limits on the computer used become a limiting factor even at moderate image sizes (e.g., 512 x 512). Thus a space domain filter must be used to filter the total SAR image, and one filter example was thus generated for the Salisbury site. Choice of a filter was dictated by cost considerations as well as image properties. A 2 x 2 equal weight averaging filter has a frequency response with the first zero crossing at a frequency of 1/2D where D is the sample interval. Power attenuation at the 50% of maximum point is 50% or - 3db at 1/4D. Higher order filters would increase attenuation; however, it was decided that within the cost constraints of the project it would be most useful to run the 2 x 2 averaging filter on the total SAR image as a demonstration. The 2 x 2 filtered image is presented at the right of Figure 60 for the Salisbury site. Slightly greater contrast is observed in the filtered image but the general appearance is not significantly changed. It was decided from this test that filtering would not significantly improve the observers ability to recognize control points in the image. The signal-to-noise ratio of this data is extremely low and filtering evidently does not help this situation. With imagery having better signal-to-noise ratio feature visibility should be better and the need for filtering less.

Correcting Geometric Errors in SAR Imagery

Correction of geometric distortions in SAR imagery to achieve registration with a reference requires generation of a geometric mapping function for each point in the digital image. The function can be based on some systematic model or can use a general polynomial. The functions need not be based on any physical quantity. Both approaches are investigated in this study. A systematic approach developed by Goodyear Aerospace was used on three test data sets (Reference 3). General polynomials were also employed on the data sets.

The three data sets were all obtained from Air Force flights using the X-Band AN/APQ-102A system over the Delmarva peninsula. These are denoted data sets No. 1, 2, and 3. No. 1 was flown over Wallops Island, Virginia on August 8, 1973. Flights No. 2 and 3 were flown on August 22, 1976 over Salisbury, Maryland (No. 2) and Cambridge, Maryland (No. 3). Data set No. 2 was used in the radiometric correction experiments discussed above. Control points from each of these three data sets were obtained and used in the various models as discussed below. The reference scene used in the case of data sets No. 2. and No. 3 was the geometrically corrected Landsat scene 2579-14535 discussed in a previous section. This data was resampled to a square 25.4 meter grid and to a UTM map projection.

Modeling Approaches

Systematic Removal Based on Error Model - The systematic approach is based on an analysis of the geometric distortions to be expected in the SAR imagery. In the systematic approach, the predictable errors are identified and correction terms are generated based on geometric parameters, and so the systematic approach may be useful whenever the geometric distortions can be modeled. The algorithms for removal of these geometric distortions are derived from an understanding of the total radar imaging system from the radar instrument and its aircraft (or spacecraft) platform, through data taking and processing to an interpretable image. The systematic approach is considered here because of its demonstrated usefulness, its simplicity, and its potential for minimizing computer processing time.

In particular, the images from the AN/APQ-102A used in the examples described in this report exhibit two predictable types of geometric distortion and an orientation problem. The predictable distortions are skew and unequal scaling in the along track and cross track directions. Skew (nonorthogonality) of the axes can be caused by radar antenna or correlator slit misalignment. Differential scaling is a result of the separate scaling mechanism involved in the range and track directions. In addition, in general, the digitized SAR image has a grid which has a different orientation than the image or map to which the SAR image is to be registered. If a specific location is to be identified by a common line and column number for all of the registered images, then the two distortions and the orientation problem must be corrected. Because these corrections are commonly needed these algorithms are likely to be applicable to the imagery of other radar systems. This systematic approach does not correct non-linearities in the SAR image that might, for example, be caused by SAR platform perturbations and terrain height variations. If the systematic approach used at Wallops is to be employed to register digitized SAR imagery to a map or to Landsat/map image, then the steps are:

1. The complete SAR image and the corresponding Landsat/Map scenes are produced at reduced scale for examination of possible control point locations. The Landsat images examined may include any or all of its spectral bands. A combined

LANDSAT image with its areas classified according to terrain and vegetation may also be used. Image reconstruction alternatives are discussed in the Appendix. The control points selected are usually cultural features since they have been found to be more reliably identifiable.

2. Images at full scale of areas surrounding tentative control points are produced for final evaluation. This allows accuracy to the nearest pixel.
3. The coordinates of control points from both the SAR image and the Landsat/Map image to which the SAR is to be registered are input into the systematic approach algorithm. Outputs of the routine are rotation angle, range scale, track scale and skew angle corrections, along with tables and plots of the residual errors for each control point. The following steps are carried out by the program:
   a. The Landsat/map coordinates of the control points are preliminarily aligned with the SAR image coordinates. This is done by determining the relative orientation of two user designated control points in both the SAR image and Landsat/map image frames and then rotating the SAR coordinate system to coincide.
   b. A least-squares fit between SAR image and Landsat/map range control point coordinates is performed using the preliminary alignment. The range errors remaining after the fit are computed and the linear correlation coefficient between range errors and track coordinates is determined (if a non-zero coefficient exists, it indicates a residual misalignment). The SAR coordinates are then rotated so as to make the coefficient zero. This process prevents individual control point errors from introducing substantial alignment errors.
   c. A least squares fit between the range coordinates of the SAR image and control point coordinates on the Landsat/map image is computed and residual errors determined (at Landsat/map scale).
   d. Track coordinates of the SAR image are scaled to the Landsat/map image via a least-squares fit and average error determined. Skew is then introduced into the SAR coordinate system via two equations:

$$Y' = Y \qquad (1)$$

$$X'(I) = X(I) + A \cdot Y(I) \qquad (2)$$

(X,Y) - original image coordinates

(X',Y') - skewed coordinates

A - tangent of skew angle

   The skew angle is varied (in sign and magnitude) until track errors are minimized (as measured by successive least-squares fits).
   e. Several types of analyses are then performed by the program in order to demonstrate the relative contribution of various error sources. In each of

them, the residual error variance and individual point errors (Landsat/map scale) are computed and displayed for examination after various types of image correction are introduced. The four types of correction are:

(1) A magnification equal to the average (range and track) scaling difference between SAR and Landsat/map coordinates.

(2) Differential scale correction.

(3) Magnification plus skew correction.

(4) Differential scale and skew correction.

The SAR image range scale, track scale and skew are displayed.

4. Re-evaluation of the control points takes place using two methods: manual and automatic.

Manual - Most important in this procedure is the final plot produced by the Systematic Approach Program for evaluation by the investigator. It is a layout of the control points in positions corresponding to their relative positions in the image. A vector is attached to each point with the magnitude and direction of its residual error. Its magnitude may be expanded in relation to the scale with which the control points are laid out. If a control point's error vector is unique in magnitude or direction with respect to its neighbors, it is suspected of being inaccurate and it is re-examined in the SAR and Landsat/map images. If the vectors in a certain region point in the same direction, this portion may be registered separately for increased accuracy. For example, Data Set #2 was divided into separate sections on this premise. The improvement in results is shown in Table 12 .

Table 12 . RESIDUAL ERRORS (STANDARD DEVIATIONS) AFTER REGISTRATION OF DATA SET #2 WHEN USING 25.4 METER PIXELS

Section Number	Control Points	Area (km^2)	Residual Errors Along Track	Residual Errors Across Track
1 of 1	38	1650	10.834	4.692
1 of 2	25	680	5.837	3.664
2 of 2	13	970	5.929	3.828
1 of 4	9	290	1.517	1.272
2 of 4	16	390	5.412	4.394
3 of 4	10	730	2.990	2.663
4 of 4	3	240	.0031	.7334

If magnitudes are large while directions of residual error vectors are without sequence or order, it indicates poor control point selection. In all cases, if the desired accuracy has not been obtained, the control points are re-evaluated and re-input into the program.

Automatic - A modified program may be used which includes automatic control point rejection, in which case it processes the control points and determines if the overall residual error is smaller than an acceptable maximum. If so, it stops; if not, it deletes the control point with the largest error and re-processes the new subset of the original control points. If at the end of the program, a sufficient number of well distributed control points do not remain, such that all areas of the image cannot be considered adequately registered, manual re-evaluation of the rejected control points is necessary. The program can then be re-run.

Polynomial Based Error Removal - Use of first thru fifth order polynomials was investigated using elements of the Laboratory for Applications of Remote Sensing image registration system and subroutines from standard scientific subroutine packages. IBM FSD also utilized third and fifth order general polynomials to describe the SAR image distortions. The polynomials of order higher than two were fit to control points from data set No. 2 and No. 3 only. Too few points were available from No. 1 to permit up to fifth order tests.

The mapping function evaluation carried out was based on the SAR imagery from two flights over Maryland's Eastern Shore on August 22, 1976. The data set names are "Salisbury" and "Cambridge". Control points were chosen throughout the scenes and conjugate points were located in the Landsat images obtained August 23, 1976 and geometrically corrected by IBM FSD. The analysis for the two strips will be discussed separately.

1. Salisbury, Maryland SAR Data (Data Set No. 2) - The Salisbury image covered nominally a 16 km (10 mi) wide (across track) by 128 km (80 mi) along track area and 47 control points were chosen over the full width and over a 55 km (34 mi) segment encompassing Salisbury, Maryland. The control points were fitted with two dimensional first through fifth order polynomials using least square approximation. Large residual errors were observed and certain points were judged as unacceptable and were removed. The fits were recomputed and considerable error remained especially in the along track case. The results of the curve fitting analysis is shown in Table 13.

The simplest approximation function tested was the Affine model which has the form:

$$v = H_1 + H_2x + H_3y$$

$$u = V_1 + V_2 + V_3y$$

Table 13. LEAST SQUARES POLYNOMIAL APPROXIMATION ERRORS FOR SALISBURY SITE

Type of Polynomial	Section	Along Track RMS Error (Pixels)	Across Track RMS Error (Pixels)	Comments
Affine	Entire	12.5	10.4	Before analysis and deletion of questionable points. (47 initial points)
Affine	Entire	10.4	3.5	After analysis and deletion of points. (34 points remaining)
Affine	Top Half	7.4	2.9	Sections overlap to produce a spline like fit.
Affine	Bottom Half	12.4	3.3	
Biquadratic	Entire	11.9	9.1	Before analysis and deletion of questionable points.
Biquadratic	Entire	9.3	3.2	After analysis and deletion of questionable points.
Biquadratic	Top Fourth	1.4	2.1	Sections overlap to produce a spline like fit.
Biquadratic	2nd Fourth	1.6	1.2	
Biquadratic	3rd Fourth	1.9	1.7	
Biquadratic	Bottom Fourth	1.8	1.5	
Bicubic	Entire	8.7	8.2	Before deletion.
Bicubic	Entire	4.6	2.0	After deletion.
Bicubic	Top Half	2.5	1.6	Sections overlap.
Bicubic	Bottom Half	2.5	1.7	
Biquartic	Entire	7.0	6.9	Before analysis and deletion of bad points.
Biquartic	Entire	3.6	1.8	After deletion.
Biquintic with 3 terms removed		2.5	2.5	34 points used.

This pair of equations can represent rotation, skew, and scale for the x axis and scale for the y axis. The fit errors for this model are presented first in Table 13. A very large RMS error was observed for the initial 47 points. Each error was inspected and the point was judged valid or misplaced by inspection of the imagery and making a subjective decision. After this process 34 points remained. The fit error was reduced to 10.44 pixels in the along track dimension which was still unacceptably large. The scene was then broken up into along track segments and subsets of the control points used in the model fits. In this case the smallest along track error observed was 7.4 samples, still an unacceptable error.

A truncated biquadratic model was used next and this equation pair is of the form:

$$v = H_1 + H_2x + H_3y + H_4xy + H_5x^2 + H_6y^2$$

$$u = V_1 + V_2x + V_3y + V_4xy + V_5x^2 + V_6y^2$$

The results for one equation for the entire area were not much better than for the Affine case. The site was then divided into four segments and fits made for each segment. In this case the largest error was 1.9 cells in the along track direction, a reasonable result.

To explore the error characteristics of higher order fits third, fourth, and fifth order polynomials were fit to the 34 high confidence control points. A standard least squares technique was used to obtain the polynomials. The higher order cases are conveniently expressed in terms of matrix notation. The fifth order case will be outlined here.

Let:

(y_i, x_i) = horizontal and vertical pixel coordinates of i^{th} control point in corrected Landsat space.

(v_i, u_i) = horizontal and vertical pixel coordinates of i^{th} control point in uncorrected SAR space.

$$T = \begin{bmatrix} v_1 & u_1 \\ \vdots & \vdots \\ v_n & u_n \end{bmatrix}$$

$$W = \begin{bmatrix} f_1(y_1,x_1) & \cdots & f_{21}(y_1,x_1) \\ & \vdots & \\ f_1(y_n,x_n) & \cdots & f_{21}(y_n,x_n) \end{bmatrix}$$

$$R = \begin{bmatrix} H_1 & V_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ H_{21} & V_{21} \end{bmatrix}$$

The $f_i(y_j,x_j)$ are the products of powers of x and y.

The columns of R will represent the coefficients of two polynomials. That is,

$$\begin{aligned} H(y,x) = {} & H_1 + H_2x + H_3x^2 + H_yx^3 + H_5x^4 + H_6x^5 + \\ & H_7y + H_8yx + H_9yx^2 + H_{10}yx^3 + H_{11}yx^4 + \\ & H_{12}y^2 + H_{13}y^2x + H_{14}y^2x^2 + H_{15}y^2x^3 + \\ & H_{16}y^3 + H_{17}y^3x + H_{18}y^3x^2 + \\ & H_{19}y^4 + H_{20}y^4x + \\ & H_{21}y^5 \\ = {} & \sum_{i=1}^{21} H_i f_i(y,x) \end{aligned}$$

$$V(y,x) = \sum_{i=1}^{21} V_i f_i(y,x)$$

The polynomials H and V approximate the mapping

$$(y_i,x_i) \longrightarrow (v_i,u_i) \quad \text{for } i=1,2,\ldots,n$$

in the least squares sense. That is,

$$\left.\begin{aligned} H(y_i,x_i) &\approx v_i \\ V(y_i,x_i) &\approx u_i \end{aligned}\right\} \text{ for } i=1,2,\ldots,n$$

The coefficients of H and V are found by solving the matrix equation T=WR for R. The solution is:

$$R = (W^T W)^{-1} W^T T$$

The equations given above are for full, fifth-degree, bivariate polynomials. All other orders use subsets of the terms listed in the above case. In each case, the form of the polynomials was determined empirically. Several polynomials were found by performing the least squares fit, and the resulting residual errors were computed. The ideal situation in doing a direct fit to control points is to have a highly overdetermined system and to have very low residual errors. This would indicate that the polynomials were good models of the geometric distortions. This did not happen for the two images in question.

The bi-cubic polynomial for the entire area produced a large error of 8.7 pixels and when split into two sections the error was still a large 2.5. The bi-cubic polynomial has 16 terms. The fourth order polynomial produced a maximum error of 7.0 and the fifth order had an error of 2.5 pixels both over the entire area. Smaller subsets were not used since there are large numbers of terms in the fourth and fifth order cases and there are only 34 control points. Three terms were removed from the fifth order polynomial to enable a nonsingular solution.

It was evident from these results that the geometric distortion in the SAR imagery with respect to the rectified Landsat was severe and single polynomials of high order would not represent the distortion adequately. To get a better look at the spatial relationship of the errors the position of each control point was plotted with vector indicating the fit error and the line and column errors were also plotted in bar graph form along the borders. This plot is shown in Figure 63. The along track (line) error demonstrates an oscillatory nature with at least four peaks over the along track span. The cause of such an oscillation is not known; however, film transport speed variations are a possible cause.

The favorable results using the segmented approach to the polynomial approximation suggested that this was an attractive approach to pursue since the order of fit in a particular segment is limited. However, it was judged that the nature of the distortions in the Salisbury data set was abnormal and more typical better quality data sets would likely not have such severe distortions.

2. Cambridge, Maryland SAR Data (No. 3) - A SAR flight over nearby Cambridge, Maryland produced a data set which was more uniform in geometry than Data Set No. 2 and lower curve fit errors were observed. Table 14 contains the RMS errors for affine through

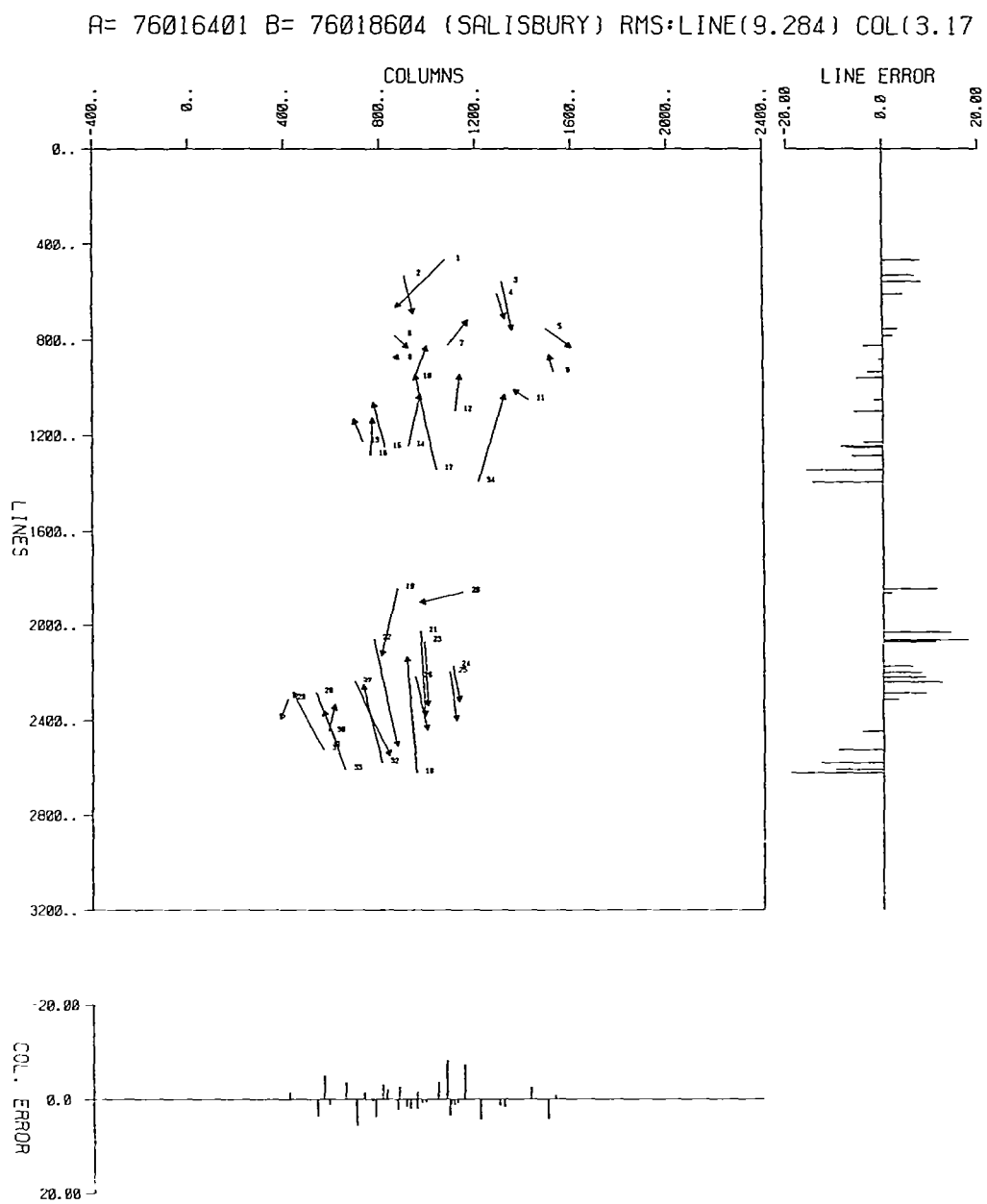


Figure 63. Vector plot of residual control points fit errors for Salisbury site

fifth order polynomial least squares fits to 47 control points. The errors for the bi-quadratic fit are plotted in Figure 67 with the line and column errors graphed along the edges of the vector plot. The LARS BIQUAD program deletes checkpoints until an error criterion is met. In the case of the Cambridge data thirty-one points were deleted and the line and column RMS errors quoted are the sixteen remaining points. All other results in Table 14 are for a fit to 37 points. Also, the BIQUAD program uses a polynomial with three quadratic terms removed whereas all other cases use all terms appropriate to the order indicated.

Table 14. RESIDUAL ERRORS (STANDARD DEVIATIONS) FOR DATA SET #3 (CAMBRIDGE) WHEN USING 25.4 METER PIXELS

Registration Approach	Residual Errors Along Track	Residual Errors Across Track
Systematic	2.29	1.83
Affine	5.00	4.04
Biquad All Points	3.98	3.96
Biquad Reduced Set	1.14	1.18
Second Order Polynomial	4.13	4.16
Third Order Polynomial	4.28	3.48
Fourth Order Polynomial	3.55	3.67
Fifth Order Polynomial	3.82	3.38

Another point should be mentioned, the plots in Figures 63 and 64 are for all the control points considered evaluated by a polynomial fit to the reduced set accepted by the program. In the case of the Cambridge plot this means that the polynomial was fit to 16 points with a 1.441 and 1.180 line and column RMS error and the error for all 47 points was plotted using this function.

The results of these tests indicate that the systematic and BIQUAD geometric error modeling approaches provide adequate means of expressing the distortion in small SAR data sets. The performance of the higher order polynomials does not appear to warrant their use. Furthermore, the cost of actually carrying out a high order polynomial registration on the entire data set would be excessive.

3. Data Set No. 1 Evaluations - The earlier (1973) data set covered a small area and was processed at Landsat resolution rather than 25m resolution as was No. 2 and No. 3.

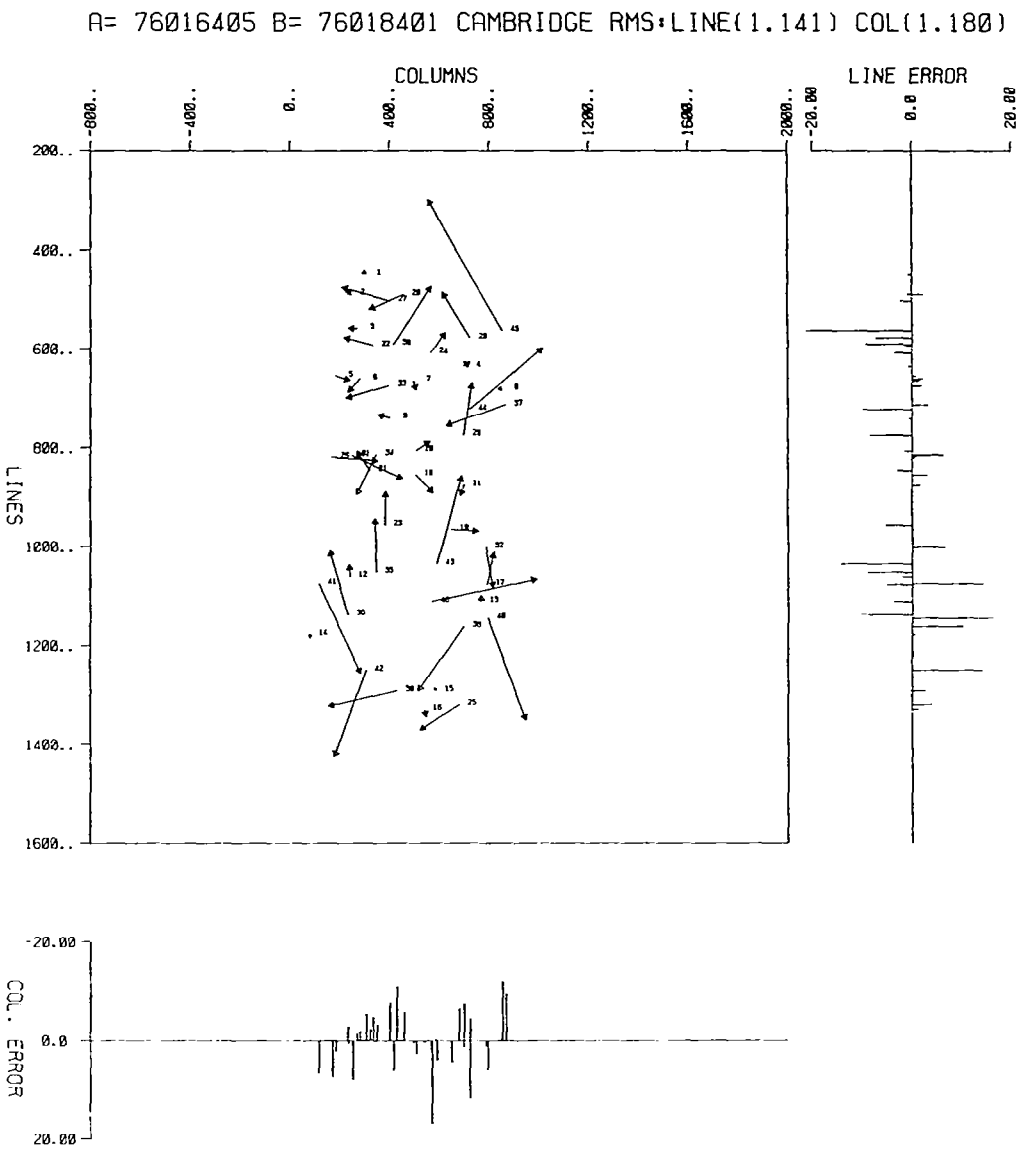


Figure 64. Vector plot of control point fit errors for Cambridge site

Table 15 contains standard deviations of the pixel differences between control points in the LANDSAT images and the control points in the SAR images after registration of Data Set No. 1 using the three different approaches. Data Set No. 1 contains 11 control points covering approximately 80 square kilometers.

Table 15. RESIDUAL ERRORS (STANDARD DEVIATIONS) FOR DATA SET NO. 1 WHEN USING 79 X 56 METER PIXELS

Registration Approach	Residual Error Along Track	Residual Error Across Track
Systematic	.4838 *	.4912
Affine	6.53 **	12.140
Biquad	4.656 **	8.171

*In reference grid
**In SAR grid

A possible explanation for the large differences in residual error magnitude may be the fact that the 11 control points used were selected from an original set of 24 using outputs from the Systematic Approach algorithm to evaluate the accuracy of each control point. Given the original selection of 24 control points, the other two approaches could conceivably find a subset of control points that would produce results equivalent to those of the Systematic Approach. That is, iterations of the systematic algorithm were used to refine the corrections for registration. The control points selected in this way may not be optimum for the other two approaches. Therefore, no conclusions should be drawn without consideration of Data Sets No. 2 and No. 3 whose residual error figures represent more accurate comparisons since the control points input were not re-evaluated with the results of a previous output.

Another point must be made regarding the results of the systematic error model. The errors quoted are with respect to the reference or output grid whereas errors for all the other methods are referenced to the grid of the image to be registered. When the scale factors of the two spaces are the same the results will be the same in either reference which was the case for Data Set No. 2. If the scale factors are different such as in the case of Data Set No's. 1 and 3 where the image to be registered (SAR) had a much smaller grid than the reference,the systematic error will be different (smaller in this case). The systematic and affine models are mathematically equivalent, but the magnitude of the errors are not necessarily identical for the two procedures even when the reference grids are identical.

Control Point Location Approaches

The basis of registration of the Landsat imagery and SAR imagery is the use of control points found in each image. Registration of Landsat imagery to a geographic coordinate grid is done by manual or automatic control point location. These approaches are discussed separately.

Manual

The data sets to be registered must first be reproduced in useful image format. There are a number of methods for reproducing digital images and those available at LARS are described with image examples in the Appendix. The Landsat image is inspected for scene objects likely to be visible in the SAR image. These can be the control points used to rectify the Landsat image to a geographic coordinate system.

The Landsat control points used may or may not be visible in the SAR image. Thus the SAR image must be inspected for existence of previously found Landsat control point locations. Once these are exhausted the SAR image is inspected for new points and the corresponding points in the Landsat image are searched for. Through such a cross inspection process a candidate set of control points is located.

Two factors come into play in the control point selection process. One is the number of points needed and the other is distribution of the points. The minimum number needed is, of course, the number of degrees of freedom (number of coefficients) in the distortion function being used. The advantage in selecting a larger number of points than the minimum is that errors in control point location are averaged and a more accurate fit results. Also, the model will generally not express the true distortions over the image and gathering a larger number of points over the scene averages the errors over the scene.

Distribution of points is important if the distortion function is to be representative of the whole scene. If corners or large regions do not contain control points it is likely that the fit in those areas will be poor.

Automatic Control Point Determination

Experimentation was carried out to determine if correlation methods could be used to define control points for registration of SAR and Landsat. In an operational situation the correlation would be done between original Landsat and SAR imagery which have different distortions and orientation. The numerical correlation process requires that the two images being correlated have the same rotational alignment and scale. This can be accomplished by performing a preliminary affine correction to one of the images before correlation. The correlator would then be employed to find any remaining small residual translational errors.

A numerical correlation processor was applied to the Salisbury and Cambridge data sets to explore the correlation potential of SAR and Landsat imagery. The correlations observed in a large number of tests were all very low (less than 50% correlation) and the tests were judged to be negative for these two cases. Further experimentation was beyond the scope of the present study; however, higher contrast SAR imagery and appropriate enhancement of the SAR and Landsat images may produce useful automatic correlation results. Thus, automatic control point finding in SAR/Landsat scene pairs is recommended for future study.

Resampling Considerations for SAR

Whenever registration of two or more digital image data sets which were sampled at different rates and over different grids is attempted the problem of resampling arises. A number of resampling schemes exist which are widely used in the image processing field. The simplest of these is called nearest neighbor (NN) resampling and is implemented by assigning the value of the nearest pixel to the location being resampled. Higher order resampling is the term applied to the use of linear, quadratic, cubic and high order polynomials in the resampling process. More advanced interpolation and filtering techniques can also be used which in some sense optimally filters the unwanted frequencies and resamples the image with minimum loss of the desired frequencies.

In the case considered here, the registration grid was chosen to be the sampling grid for the radar data which was nominally 25 meters. The Landsat data with which the SAR data is to be registered is sampled at 57 meters in the across track direction and 79 meters in the along track direction. The Landsat data thus must be resampled to 25 meters to match the SAR data. In the registration process the SAR data is geometrically transformed but since the "before" and "after" grids are both 25 meters no scale change is necessary. Thus, the problem of resampling the Landsat data was studied in detail and the assumption that nearest neighbor resampling would be adequate for the SAR data was made. The maximum position error for square grid nearest neighbor resampling is $D/\sqrt{2}$ or for this case $25/\sqrt{2} = 17.7$ meters. For the Landsat data the NN error would be much larger and the project resources were thus put into the Landsat resampling problem. Further study is needed of the effects of higher order resampling on the SAR image.

Recommendations for Control Point Identification and Geometric Modeling Procedures

The data sets available in this study had relatively low signal-to-noise ratios and extensive distortion function research could not be effectively carried out. Thus, further study into the proper form for the mapping polynomials is needed. The size of the subimage being corrected is also quite significant. The larger the SAR image, the more difficult it is to characterize the distortions by polynomials. The number of RCPs available directly affects the possible forms of the mapping polynomials. For the

investigation conducted here the Systematic and Biquadratic approaches appear to be adequate. It is recommended, however, that the systematic, and first through fifth order polynomial fitting algorithms be made available to users of an operational system. The manual control point finding procedure is recommended based on results in this study using the various image reconstruction alternatives outlined in the Appendix.

Special Aspects of SAR Satellite Imagery

In general, radar imagery from satellites will be identical to that obtained from aircraft, given equivalent collection geometry, and system parameters. In practice, it will probably be some time before satellite imagery having wide swaths and/or shallow depression angles is available because of hardware cost limitations, but there are no technology problems preventing such operation.

One specific difference between satellite and aircraft imagery will be a smaller change in depression angle across a swath because of the greater ranges involved. A more constant depression angle will facilitate interpretation of the imagery because this important variable will be essentially constant over a given image.

EVALUATION OF TECHNICAL RESULTS

Landsat Processing Evaluation

The Landsat scene used for Data Set #1 (1403-15132) was acquired on August 30, 1973 by the MSS of Landsat-1. Figure 65 is the uncorrected Band 6 of this Landsat scene. The scene was then geometrically corrected by LARS using a systematic approach that is part of a standard software package at LARS. This resulted in Landsat/Map registration errors greater than one pixel.

The Landsat-2 scene used for Data Sets #2 and #3 (2579-14535) were processed by IBM under Contract NAS6-2827. The processing performed was scene correction, i.e., the image was geometrically corrected to a Universal Transverse Mercator (UTM) map projection using geodetic control points to model scene dependent errors (attitude and altitude). Digital data at two scales were produced. First, the entire scene was processed, resulting in an image with pixels at a spacing of 50.8 meters in both the horizontal and the vertical directions. Figure 66 is a print of Channel 6 of this entire scene at the 50.8 meters pixel spacing. Second, two subimages of the scene were processed with pixels spaced at 25.4 meters in both directions. Figure 67 is a color composite of channels 4, 5, and 7 of the subimage near Salisbury, Maryland, with pixels spaced at 25.4 meters, used for Data Set #2. Figure 68 is a color composite of the subimage near Cambridge, Maryland, used for Data Set #3.

The geometric accuracy of the corrected Landsat images was evaluated by checking the errors at the locations of the geodetic control points (GCPs) that were used to correct the image. For each GCP, a nominal input-space location was found by the following transformations:

1. Mapping the UTM coordinates into output-space pixel coordinates using an elementary rotation, translation, and scale-change transformation.

2. Mapping the output-space coordinates into input-space pixel coordinates using the mapping polynomials that define the geometric transformation for the scene.

The nominal input-space locations were compared with the corresponding observed input-space locations that were found manually from the input data. The resulting errors in meters were found. The RMS of these errors, as shown in Table 16, was 77.3 meters.

The corrected digital data was recorded on film. The resulting photographic image products were of good quality, as can be seen in Figures 66, 67 and 68.

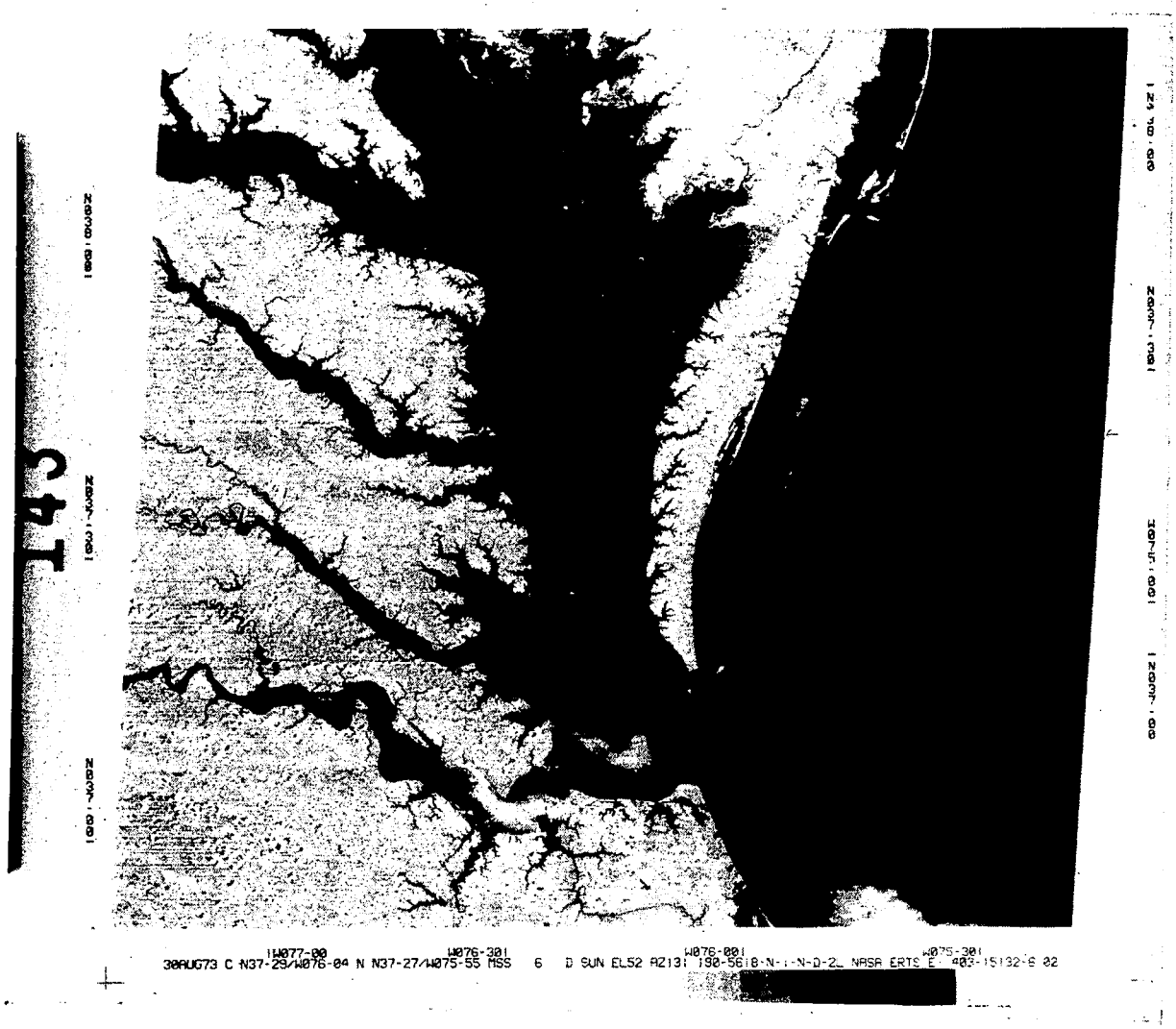


Figure 65. Uncorrected Landsat scene used for Data Set 1

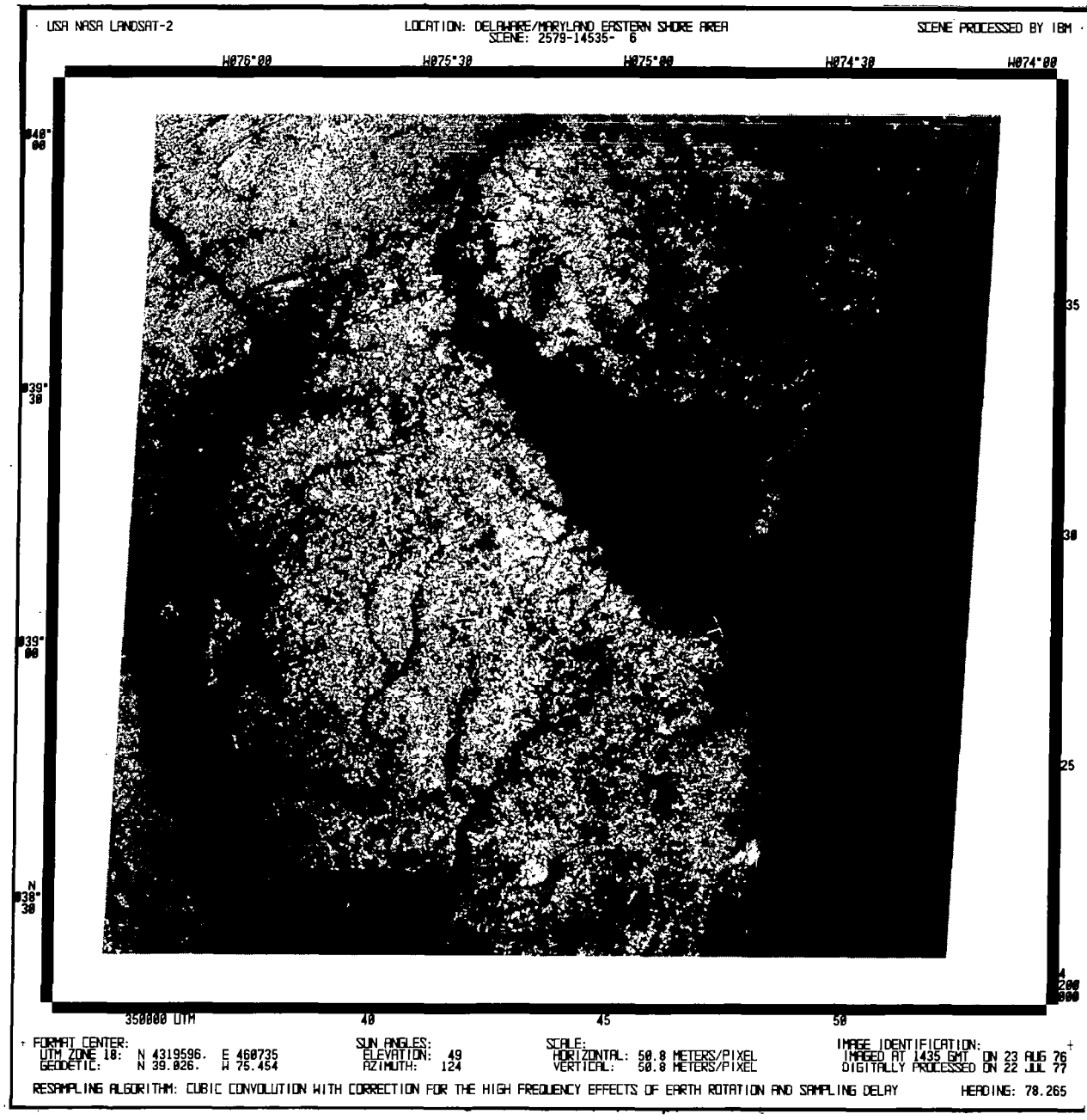


Figure 66. Corrected Landsat scene used for Data Sets 2 and 3

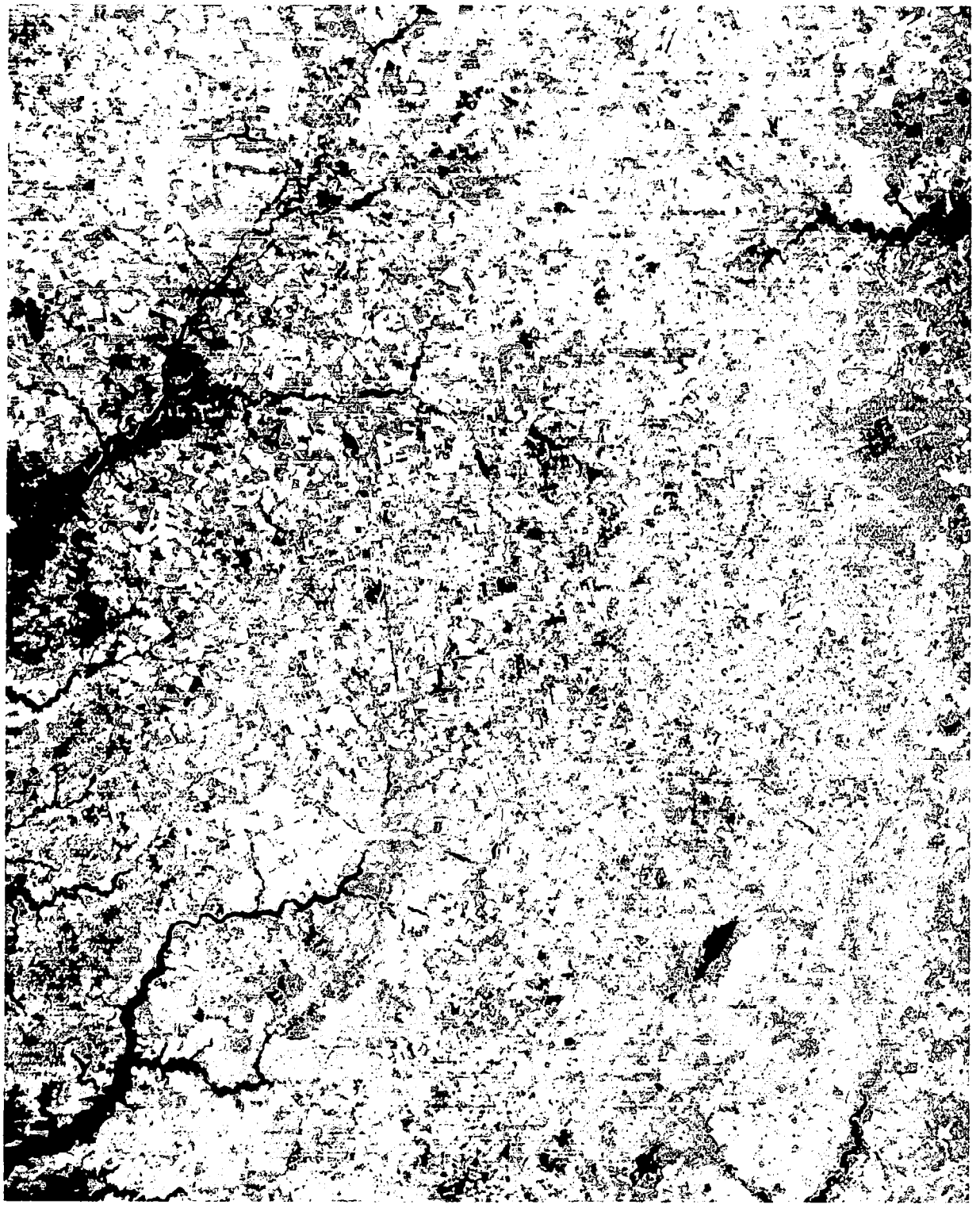


Figure 67. Color composite of corrected Landsat 2 scene of Salisbury, Md. used for Data Set 2

Figure 68. False color IR reproduction of corrected Landsat data for the Cambridge, Maryland site for Data Set 3

Table 16. RESULTS OF ACCURACY CHECK

GCP	NORTHING	EASTING	ZONE	YOBS	XOBS	YNOM	XNOM	DELY	DELX	RSSM
1	4368380	399200	18	361.00	729.00	359.69	730.32	1.3081	-1.3200	128.517
2	4319160	370820	18	91.00	1402.00	90.57	1402.00	0.4348	-0.0004	24.843
3	4284533	450858	18	1618.00	1623.00	1618.81	1622.84	-0.8133	0.1637	48.244
4	4281320	465530	18	1883.00	1625.00	1883.10	1625.06	-0.0990	-0.0551	7.145
5	4265817	477486	18	2155.00	1784.00	2156.48	1784.25	-1.4757	-0.2456	86.537
6	4269440	407370	18	945.00	1915.00	944.69	1917.11	0.3063	-2.1149	168.413
7	4249185	379330	18	554.00	2236.00	544.46	2236.31	-0.4646	-0.3072	36.014
8	4235348	419500	18	1303.00	2304.00	1303.69	2303.83	-0.6935	0.1690	41.826
9	4238490	420600	18	1308.00	2262.00	1308.49	2262.53	-0.4887	-0.5264	50.181
10	4243210	370290	18	425.00	2333.00	425.45	2332.68	-0.4459	0.3172	35.783
11	4241985	375720	18	524.00	2333.00	524.59	2333.89	-0.5905	-0.8859	77.853
12	4390075	399120	18	261.00	465.00	260.61	465.84	0.3899	-0.8435	70.425
13	4382940	408410	18	454.00	529.00	453.25	529.51	0.7472	-0.5134	58.959
14	4377045	413810	18	573.00	586.00	572.66	587.82	0.3373	-1.8218	145.569
15	4350880	387920	18	244.00	971.00	243.51	972.12	0.4946	-1.1209	93.167
16	4378540	449910	18	1185.00	478.00	1183.88	478.70	1.1172	-0.6984	84.470
17	4381505	451450	18	1196.00	438.00	1197.01	438.67	-1.0054	-0.6745	78.452
18	4372495	453540	18	1273.00	543.00	1272.57	543.26	0.4295	-0.2598	32.022
19	4367805	453865	18	1298.00	599.00	1298.92	599.61	-0.9172	-0.6115	71.356
20	4379950	425480	18	760.00	523.00	760.06	522.99	-0.0573	0.0066	3.314
21	4373880	422460	18	736.00	604.00	735.36	604.63	0.6420	-0.6303	61.952
22	4357310	493810	18	2026.00	626.00	2025.99	626.98	0.0140	-0.9799	77.609
23	4319340	516745	18	2589.00	1031.00	2588.68	1031.95	0.3244	-0.9549	77.866
24	4317520	511740	18	2510.00	1067.00	2510.85	1066.76	-0.8480	0.2375	51.979
25	4273070	494460	18	2414.00	1652.00	2414.46	1652.81	-0.4622	-0.8053	69.032
					MEAN OF EACH COLUMN:			-0.0726	-0.5790	67.261
					RMS OF RSSM:					77.301

SAR Image Formation Evaluation

Data Set No. 1

SAR imagery was acquired of portions of the Delmarva Peninsula by an United States Air Force RF-4 aircraft based at Shaw Air Force Base, South Carolina, and Landsat imagery was acquired of the same date. The SAR was a Goodyear Aerospace AN/APQ-102A (X-band). The selection of this particular SAR scene was based on experience with a misclassification problem with urban areas and certain beach and field areas which resulted when Landsat imagery only was used (see Reference 4). An appropriate portion of the SAR data film was optically correlated by Goodyear and the resultant imagery was provided to Wallops as a film transparency. This image was then scanned on the Wallops microdensitometer at a scanning aperture of fifty micrometers. Since the scale of the image was 1:800,000, the microdensitometer produced a digital image with a pixel size approximately equal to a square with forty meter sides. Scratches are noticeable in the SAR imagery (see Figure 69).

Figure 69. SAR image of Wallops Island - Assateague area for Data Set 1

Data Set No. 2

SAR imagery was acquired of portion of the Delmarva Peninsula by an United States Air Force RF-4 aircraft based at Shaw Air Force Base, South Carolina. The SAR imagery was acquired on August 22, 1976 by a Goodyear Aerospace AN/APQ-102A. SAR coverage included an area extending from Salisbury, Maryland, to Georgetown, Delaware. This scene was selected because it included a mixture of urban and agricultural areas, and because considerable ground truth was available. The doppler phase history was delivered on film to Wallops by the USAF. The USAF also formed an image from the phase history and delivered this image to Wallops (see Figure 70). The same phase history was sent to Goodyear Aerospace who formed and delivered to Wallops an image (see Figure 71). However, these images were recorded on film and so were limited to a dynamic range of contrast of about 30 db (density range of about 0 to 3). Therefore, the phase history was set to ERIM in order that the formed image be recorded directly on magnetic tape so as to preserve the dynamic range as well as to provide a digital image suitable for computer processing.

ERIM recorded the data at two different resolutions, 15 and 50 meters. Resolution is defined here as resolvable element pairs, so the 50 meter data means an individual pixel size of 7.5 meters. ERIM delivered the above images to Wallops on magnetic tape. The contents of the ERIM supplied tape of the 25 meter image of the Salisbury area was imaged on the Wallops filmwriter and is shown as the leftmost print of Figure 60. The ERIM supplied tape of the 7.5 meter image of the Salisbury area was also imaged on the Wallops filmwriter and is displayed as Figure 72. These images were contrast stretched in order to fill the contrast range of the filmwriter. Each digital value on these ERIM supplied tapes represents the square root of the received radar power for each pixel. The Wallops filmwriter was used to display the contents of the tapes. The display options that were used included making the film exposure proportional to the following function of the values on the digital tape: constant; square; and long of the square. This latter representation was expected to yield the best quality image (to the eye) so enhancements of specific portions of this image were executed. Noise persisted as a serious problem for all of the images produced from the digital tapes.

It is not clear why the images of Figures 60 and 72 seem to contain more noise than the images of Figures 70 and 71. A contributing cause could be that too few looks were used when the image was formed. The greater the number of looks used, the smoother the output data; although the number of looks can not be increased to the theoretical limit since aircraft motions preclude the exact alignment of each doppler phase. The spatial qualities of the image eventually begin to degrade as the number of looks increase. For the 7.5 meter pixel image a single look only may be necessary, but for the 25 meter pixel image multiple looks could have been used.

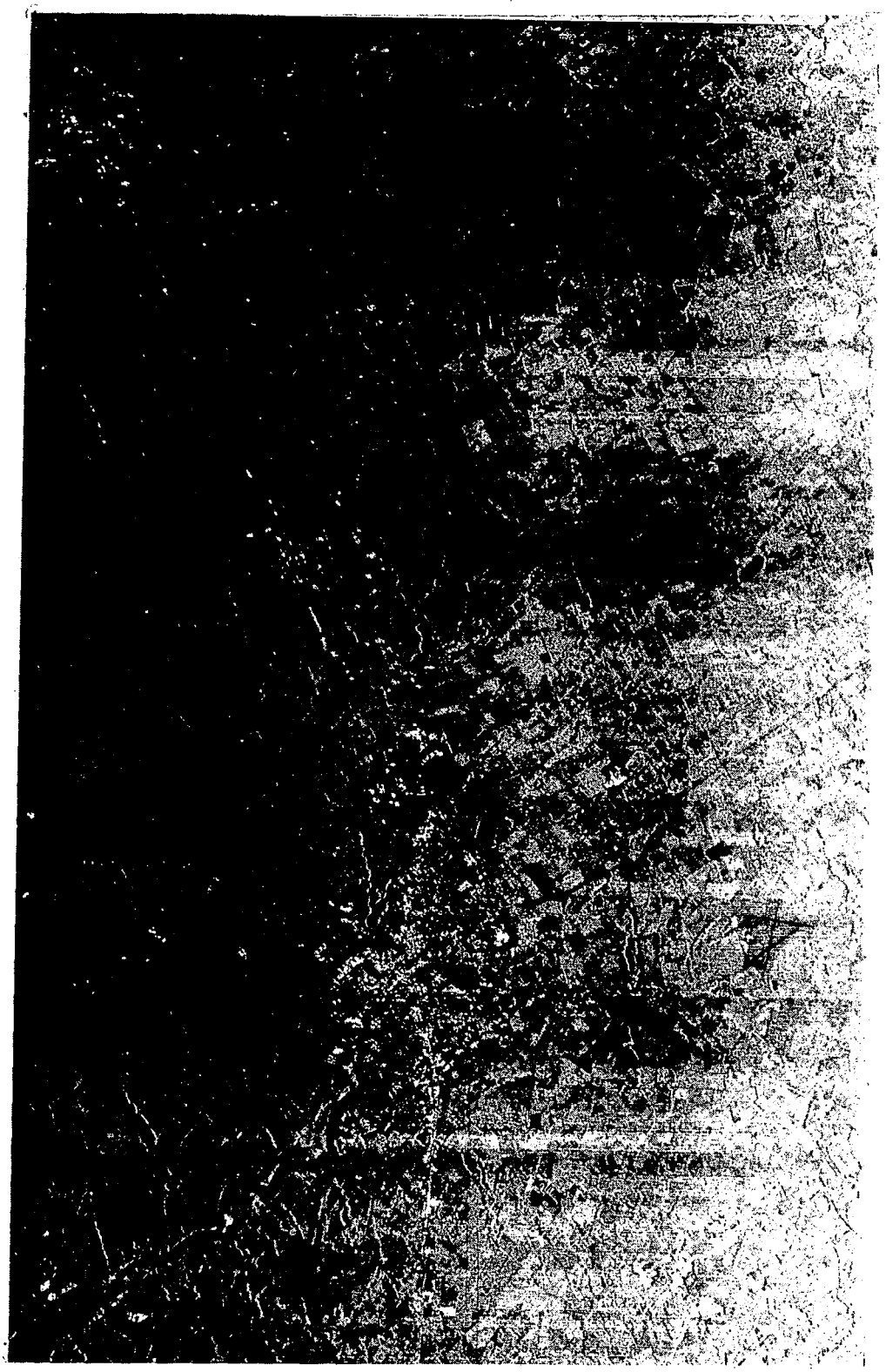


Figure 70. SAR image formed by Air Force of Salisbury area

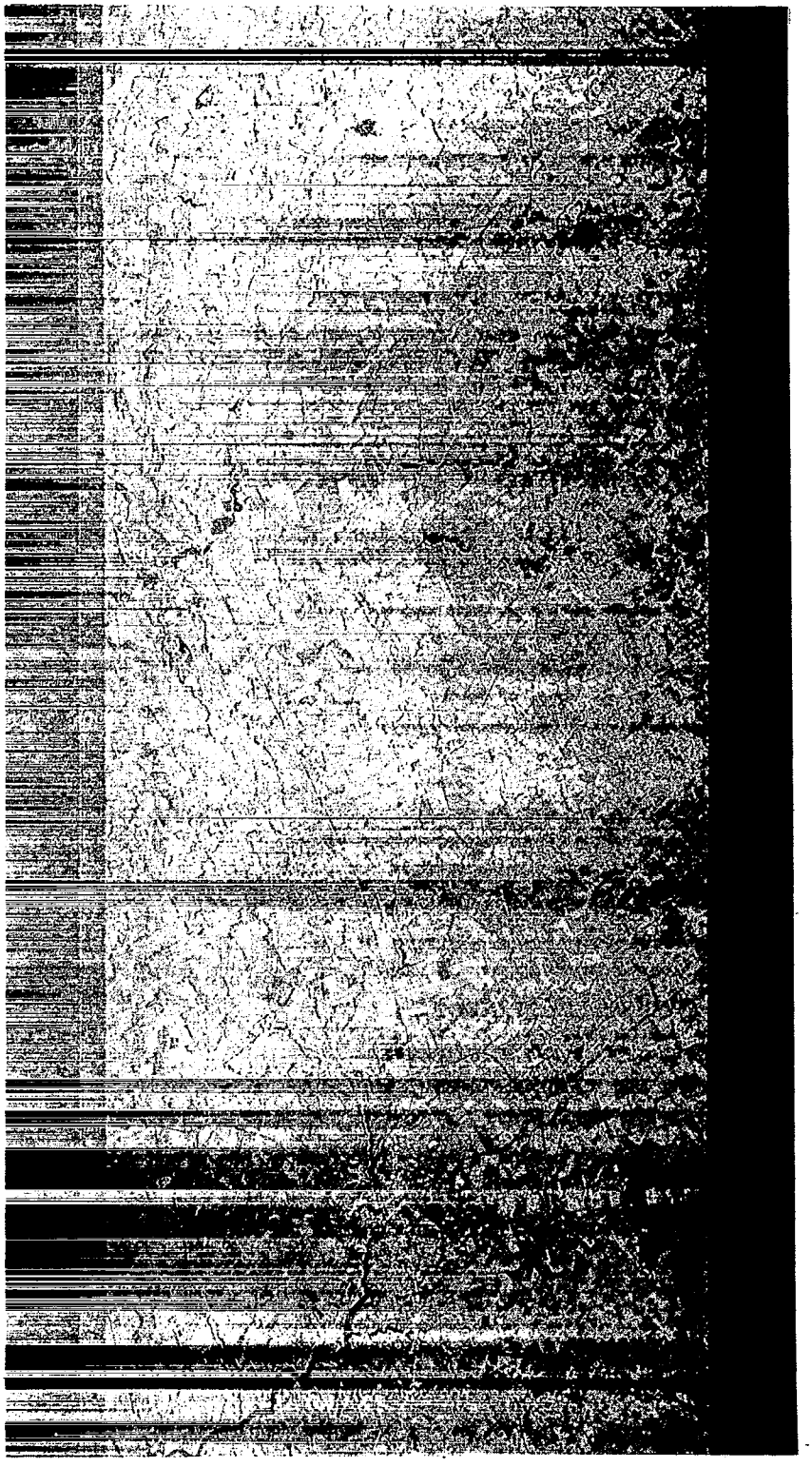


Figure 71. SAR image formed by Goodyear of Salisbury area

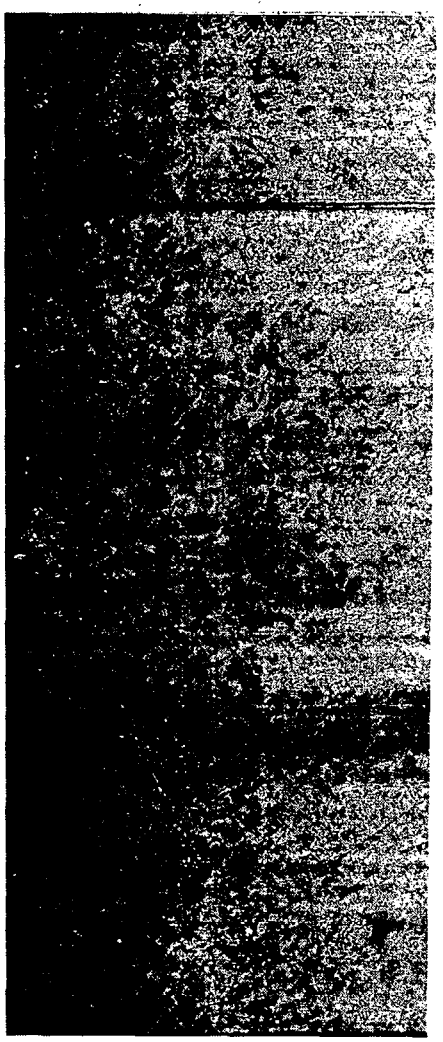


Figure 72. SAR image formed by ERIM of Salisbury area

Data Set No. 3

Optical Radar Correlation - This is an example of good quality AN/APQ-102A imagery, for both resolution and dynamic range. However, some along track lines are on the imagery as a result of scratches on the AF supplied doppler data film. The SAR image was formed by Goodyear Aerospace and delivered to Wallops on film.

IBM Digital Print - The digital image for data set number three was produced by scanning a black-and-white-print of a SAR image of the Cambridge, Maryland area. The scanning was performed on the IBM Drum Scanner/Plotter. A nominally square aperture of size 0.004 inches by 0.004 inches was used. The pixel spacing was 0.004 inches in both the horizontal and the vertical directions.

The resulting image is shown in Figure 73. Although there is still some noise evident in the data, the quality of the image is considerably better than that of data set number two. There are several lines in the image that appear to be noise. They are probably a result of scratches in the negative used to produce the print from which the digital data was scanned.

Figure 73. SAR image for Cambridge site scanned from source print by IBM film scanner

SAR Image Processing Evaluation

Data Set No. 1

Over the area of approximately five km by eight km (3 miles by 5 miles) the systematic errors were removed from the SAR imagery (Reference 3). The SAR imagery was registered with the geometrically corrected Landsat imagery by the use of 11 control points. However, scratches in the imagery caused severe problems in selecting these control points. The distortion corrections resulted in a mean error along track of 0.38 pixels with a standard deviation of 0.69 pixels.

The Landsat and SAR imagery were at different resolutions, but the resolutions need to be identical if computer based multispectral analysis were to be performed using LARSYS. Therefore, the SAR image was degraded to that of Landsat using a nearest neighbor approach.

The registered imagery was analyzed on several facilities in order to identify the impact of the SAR channel (Reference 5). For example, Figure 74 is a color composite of two corrected Landsat MSS bands and the corrected SAR image as displayed on the Goddard Image 100. Figure 75, also generated on the Goddard Image 100, is a Landsat color composite merged with the high returns from the SAR channel. Another facility used for analysis was a remote terminal to Purdue/LARS. This terminal was used to access a set of multispectral image analysis computer application program called LARSYS (Reference 6).

The merged imagery of the selected scene was clustered with LARSYS using both the four and the five channel data sets (Reference 6). Subsequent separability studies indicated that there were twenty discrete classes for the four channel Landsat scene, and twenty-six discrete classes for the five channel Landsat/SAR data. Further analysis of classification accuracies indicated that two classes associated with urban areas were separable with the five channel data set and were not separable with the four channel Landsat scene. Several other classes were attributable only to scratches in the SAR imagery and to water/land interface areas where the Landsat MSS and the SAR "saw" the coastline differently. Due to these limitations and the ability to register only limited areas of SAR imagery to Landsat imagery, attention was shifted to a second set of SAR/Landsat data.

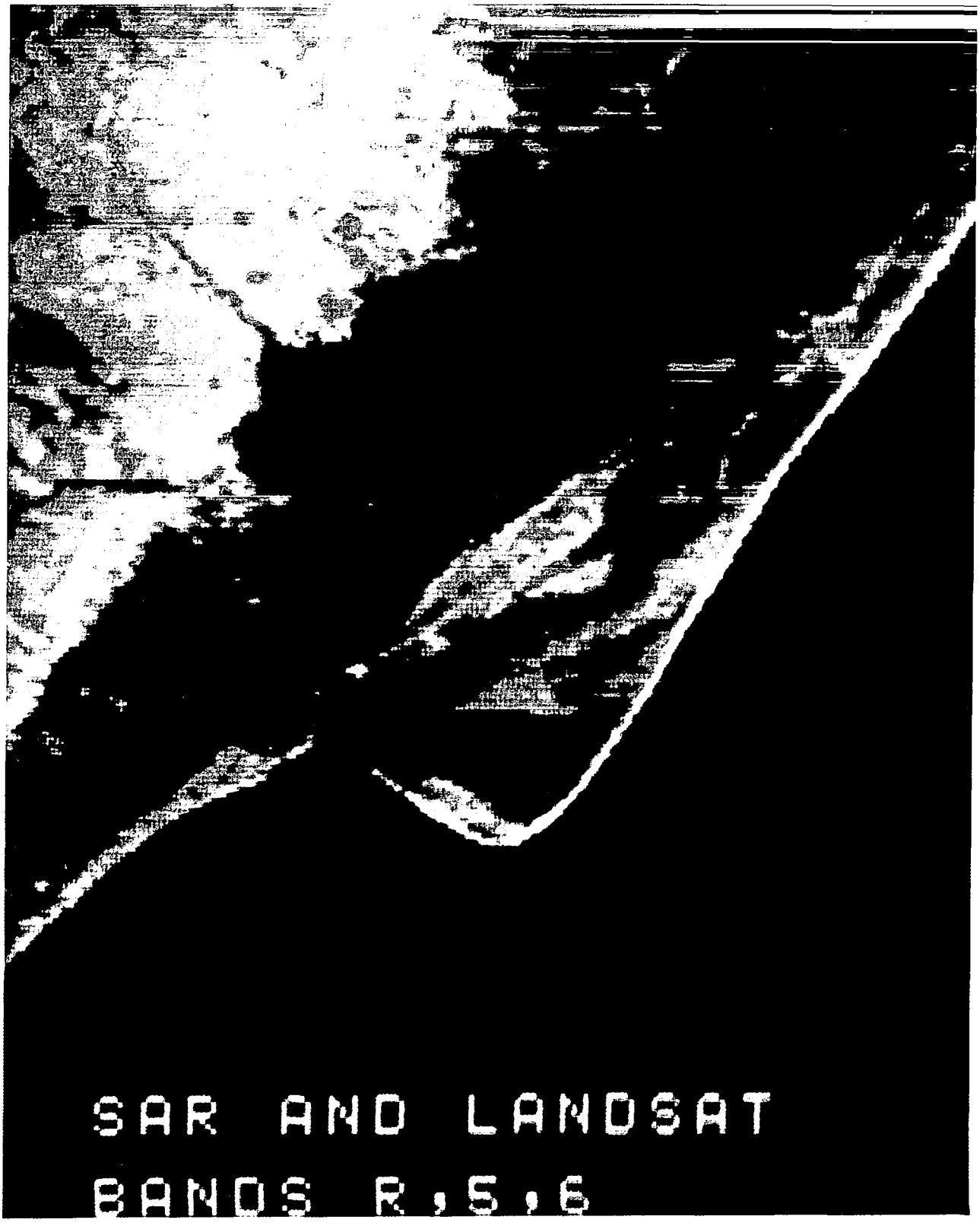


Figure 74. Color composite of one SAR and two Landsat bands from Data Set 1

Figure 75. SAR high returns registered to Landsat color composite

Data Set No. 2

If the SAR image of Data Set #2 was to be suitable for computer based analysis, then the radiometric quality needed improvement since the imagery was shaded in the across track direction and banded in the along track direction, as well as containing a high frequency salt and pepper effect. The shading was due to antenna gain while the banding was caused by the Automatic Gain Control of the SAR. The shading and banding were corrected by LARS (Figure 60). The imagery was digitally filtered in order to reduce the noise. However, when the noise was reduced by this method, the image was degraded (blurred). Figure 62 displays the original and smoothed images. The noise hindered the location of registration control points (RCP), so that IBM located only 14 while LARS located only 34. The inability to accurately locate a sufficient number of RCPs precluded using higher order algorithms for the registration since they require a large number of control points.

The spatial resolution of the SAR image for Data Set #2 was satisfactory, but due to the geometric discrepancies between the SAR and Landsat imagery, higher order algorithms were needed to adequately register the SAR to the Landsat image. Therefore, an accurate registration for Data Set #2 was not possible using a polynomial approach only.

Bivariate, linear through fifth order, least squares polynomials were tested to define the transformation between corrected and uncorrected SAR data. A biquadratic function was used by LARS and a bicubic was used by IBM to perform the registration. The RMS residual errors of the least squares fit at the RCP locations of the LARS version were as large as 9.3 pixels. The least squares fit of the IBM version had reasonable registration at the 14 RCP locations, however, some severe misregistration is evident at locations between the RCPs. For example, U.S. Route 50 east of Salisbury is clearly visible in both the SAR and the Landsat data. In the corrected SAR image, that road is displaced to the north of its proper location, as can be seen in Figure 76. RCPs above (lake) and below (airport) the misregistered road are properly registered.

Due to the lack of RCPs, it is difficult to draw any conclusions as to whether the SAR image of Data Set #2 can be corrected with any given least squares polynomial. Also, since the maximum number of terms in a least squares polynomial is directly related to the number of points being fitted, it would be difficult to obtain significant improvement in the correction accuracy of Data Set 2. The registered SAR/Landsat data set for Salisbury (LARS Run No. 76016404) that is listed in Table 2 as Data Set #2 contains the noisy image of Figure 63. It is planned to prepare a new Data Set #2 by digitizing the image of Figure 71 and registering it to the appropriate corrected Landsat image.

Table 17. RESULTS OF LEAST SQUARES FIT TO CONTROL POINTS FOR THE SALISBURY SITE

Y(I)	X(I)	YP(I)	XP(I)	YEVAL	XEVAL	DY	DX	RSS
1037.0000	1350.0000	393.0000	2058.0000	393.0143	2058.0349	-0.0143	-0.0349	0.0377
1163.0000	1847.0000	663.0000	2488.0000	662.3030	2486.4555	0.6970	1.5445	1.6945
870.0000	1853.0000	367.0000	2570.0000	366.6809	2569.6435	0.3191	0.3565	0.4785
775.0000	2068.0000	326.0000	2804.0000	327.0820	2804.3825	-1.0820	-0.3825	1.1476
1105.0000	2170.0000	685.0000	2812.0000	685.0728	2811.0775	-0.0728	0.9225	0.9254
656.0000	2203.0000	244.0000	2976.0000	243.3904	2976.2905	0.6096	-0.2905	0.6752
583.0000	2446.0000	240.0000	3265.0000	240.1656	3264.9272	-0.1656	0.0728	0.1809
653.0000	2607.0000	358.0000	3402.0000	357.9502	3402.0932	0.0498	-0.0932	0.1057
1153.0000	1861.0000	655.0000	2502.0000	655.8181	2503.6869	-0.8181	-1.6869	1.8748
1101.0000	2187.0000	685.0000	2825.0000	684.6824	2825.9888	0.3176	-0.9888	1.0385
988.0000	2150.0000	562.0000	2830.0000	562.1898	2830.0688	-0.1898	-0.0688	0.2019
917.0000	2182.0000	501.0000	2879.0000	498.9147	2879.7936	2.0853	-0.7936	2.2312
954.0000	2213.0000	542.0000	2899.0000	543.6066	2898.0484	-1.6066	0.9516	1.8673
687.0000	2246.0000	286.0000	3011.0000	286.1294	3010.5073	-0.1294	0.4927	0.5094
				MEAN OF EACH COLUMN:		-0.0000	0.0000	0.9263

Figure 76. Registered Landsat/SAR data for Salisbury site

Data Set No. 3

The same procedure as was used on Data Set No. 2 was used on Data Set No. 3, an image containing the Cambridge, Maryland area. However, 43 RCPs and fifth-degree polynomials were used for the least squares fitting process by IBM and 47 points with biquadratic polynomials were used by LARS. The results from the least squares fit are shown in Table 18 for the IBM results. The mean RSS error of about 5 pixels were obtained for the IBM version and the LARS biquad results for all points was approximately 4 pixels in the SAR grid. This is quite acceptable, since the pixels in the Cambridge image are much smaller (about 10 meters) than were those in the Salisbury image.

The resulting SAR image that was produced from second and fifth-degree polynomials was suitably registered to the Landsat image. As can be seen in Figure 80, the only misregistered region appears to be the southern bank of the Choptank River (at the lower left of the SAR image). This error can probably be corrected by careful refinement of RCP locations in the lower left region of the SAR image.

It can also be seen in Figure 80 that the SAR data adds information that was not present in the Landsat data. For example, detailed information in the Cambridge urban area is present in the SAR data, but not in the Landsat data.

The results from the experiment on Data Set No. 3 indicate that registration of SAR data with Landsat MSS is both feasible (from a technical viewpoint) and useful (from an information-content viewpoint). Multi-sensor temporal registration is likely to be a very desirable process in the future. It will enable users to make use of varied sources of information to solve their problems.

Table 18. RESULTS OF LEAST SQUARES FIT TO CONTROL POINTS FOR THE CAMBRIDGE SITE

Y(I)	X(I)	YP(I)	XP(I)	YEVAL	XEVAL	DY	DX	RSS
359.0000	236.0000	406.0000	68.0000	400.4535	66.2172	5.5465	1.7828	5.8260
339.0000	261.0000	358.0000	146.0000	361.3446	145.6428	-3.3446	0.3572	3.3636
486.0000	263.0000	717.0000	70.0000	718.2549	63.8528	-1.2549	6.1472	6.2740
531.0000	270.0000	830.0000	54.0000	830.4441	58.6436	-0.4441	-4.6436	4.6648
575.0000	316.0000	957.0000	156.0000	962.1626	159.0121	-5.1626	-3.0121	5.9771
633.0000	345.0000	1120.0000	203.0000	1118.7914	206.3516	1.2086	-3.3516	3.5628
702.0000	318.0000	1271.0000	97.0000	1268.7877	97.0309	2.2123	-0.0309	2.2125
233.0000	289.0000	114.0000	293.0000	115.5594	294.7228	-1.5594	-1.7228	2.3238
339.0000	378.0000	420.0000	461.0000	420.3333	468.8501	-0.3333	-7.8501	7.8571
387.0000	445.0000	579.0000	632.0000	577.9556	628.9808	1.0444	3.0192	3.1947
856.0000	369.0000	1696.0000	150.0000	1696.0116	149.0355	-0.0116	0.9645	0.9646
834.0000	423.0000	1659.0000	309.0000	1660.4976	304.5349	-1.4976	4.4651	4.7095
852.0000	483.0000	1731.0000	448.0000	1734.7679	458.7640	-3.7679	-10.7640	11.4044
869.0000	508.0000	1795.0000	523.0000	1791.0800	518.3597	3.9200	4.6403	6.0744
653.0000	434.0000	1222.0000	444.0000	1217.4345	441.0422	4.5655	2.9578	5.4399
538.0000	450.0000	951.0000	562.0000	949.5650	553.9528	1.4350	8.0472	8.1741
606.0000	544.0000	1164.0000	779.0000	1166.1527	778.1531	-2.1527	0.8469	2.3133
385.0000	528.0000	621.0000	866.0000	623.7306	864.9073	-2.7306	1.0927	2.9411
176.0000	509.0000	102.0000	945.0000	97.0237	937.4590	4.9763	7.5410	9.0350
162.0000	547.0000	80.0000	1051.0000	83.2893	1052.3360	-3.2893	-1.3360	3.5502

Table 18 (continued)

Y(I)	X(I)	YP(I)	XP(I)	YEVAL	XEVAL	DY	DX	RSS
351.0000	613.0000	593.0000	1128.0000	591.7437	1124.4698	1.2563	3.5302	3.7471
457.0000	617.0000	857.0000	1060.0000	850.0797	1072.1978	6.9203	-12.1978	14.0242
608.0000	654.000	1229.0000	1083.0000	1231.7469	1084.6831	-2.7469	-1.6831	3.2216
623.0000	685.0000	1281.0000	1165.0000	1284.2207	1161.9511	-3.2207	3.0489	4.4349
138.0000	625.0000	61.0000	1283.0000	62.3325	1286.7827	-1.3325	-3.7827	4.0105
153.0000	717.0000	154.0000	1541.0000	151.3820	1540.6294	2.6180	0.3706	2.6441
363.0000	755.0000	691.0000	1513.0000	697.7959	1515.5991	-6.7959	-2.5991	7.2759
346.0000	855.0000	713.0000	1822.0000	707.5786	1807.2257	5.4214	14.7743	15.7376
250.0000	781.0000	431.0000	1660.0000	435.2218	1660.9041	-4.2218	-0.9041	4.3175
284.0000	809.0000	534.0000	1713.0000	533.1318	1717.9198	0.8682	-4.9198	4.9959
782.0000	991.0000	1842.0000	1920.0000	1842.3322	1918.9094	-0.3322	1.0906	1.1400
716.0000	1011.0000	1701.0000	2016.0000	1699.6569	2018.7516	1.3431	-2.7516	3.0619
558.0000	898.0000	1257.0000	1795.0000	1249.9346	1797.4515	7.0654	-2.4515	7.4786
483.0000	899.0000	1062.0000	1846.0000	1065.9349	1845.4864	-3.9349	0.5136	3.9682
538.0000	987.0000	1251.0000	2059.0000	1250.4182	2059.4309	0.5818	-0.4309	0.7240
93.0000	1039.0000	185.0000	2480.0000	185.1883	2479.4603	-0.1883	0.5397	0.5716
337.0000	1015.0000	777.0000	2250.0000	775.8636	2254.5205	1.1364	-4.5205	4.6612
581.0000	1096.0000	1409.0000	2337.0000	1416.2309	2335.3983	-7.2309	1.6017	7.4062
645.0000	1112.0000	1581.0000	2341.0000	1582.0406	2342.6730	-1.0406	-1.6730	1.9702
644.0000	1147.0000	1601.0000	2439.0000	1594.9375	2436.3345	6.0625	2.6655	6.6226
734.0000	1110.0000	1797.0000	2280.0000	1797.2107	2281.5745	-0.2107	-1.5745	1.5885
742.0000	1116.0000	1817.0000	2292.0000	1819.1140	2292.4236	-2.1140	-0.4236	2.1560
759.0000	975.0000	1779.0000	1892.0000	1778.2642	1889.3737	0.7358	2.6263	2.7274
				MEAN OF EACH COLUMN:		-0.0000	-0.0000	4.8453

Figure 77. Registered Landsat/SAR data for Cambridge site

SAR/LANDSAT DATA MERGING SYSTEMS

A SAR/Landsat data merging system is required to provide the remote sensing community with a capability to access for analysis and further research registered data from these two sensor systems. While general registration procedures apply, special considerations must be made to cover this multisensor application. The considerations are quite broad, because the radar comes from multiple sources such as aircraft and satellites. In contrast Landsat MSS data is a relatively stabilized input data source with better known parameters. This discussion indicates that a solid, generalized-design SAR/Landsat data merging system is required to assure the remote sensing community access to this multisensor data in its most useful form.

General System Requirements

Elements of the study of the SAR/Landsat data merging system include functional requirements for both hardware and software. In addition, special user considerations and a suggested implementation plan are presented. A few comments will be made to establish a background for each of these topics.

Functional requirements refer to the basic conceptual actions which may be needed; they underlie the system of programs which make up the SAR/Landsat data merging system. Data sources and parameters are defined and input reformatting programs are provided.

The reformatted input data must be imaged, and radiometric as well as geometric corrections applied as needed. Data magnification by changing pixel spacing as well as data orientation to a true north-south grid are examples. Both SAR and Landsat data share these functional requirements.

Programs with algorithms for the actual registration of the data will be included. These programs include the ability to track and/or model errors. Control point acquirement is a problem to be addressed in both manual and automated modes. Various interpolation schemes may be utilized to aid registration.

Finally, intermixed with these is the requirement for imaging. Good usable images are currently critical in virtually all stages of the registration process. Data display needs vary from the basic line printer, to matrix line printing, to digital film writing at fairly high resolutions.

In summary, the functional requirements cover the data merging systems input processing, registration, and output processes. The choice of hardware and software are significantly impacted by these functional requirements; e.g., data orientation to a true north-south grid will need a large buffer so that elimination of this requirement would significantly reduce the hardware requirements. The size and complexity of the data

sets make a large frame computer with virtual memory desirable. Programs and programmers become less hindered by computer memory space constraints when a large main frame with virtual memory is available, and so the development of programs tends to progress faster. Magnetic tape will be the primary image data storage medium. If the computer instruction set is augmented by an array processor, then the throughput might be increased. The number of data sets which may be processed will be influenced by the flexibility and power of the computer equipment. Special hardcopy imaging hardware should also be considered as part of a total system requirement.

The functional requirements are implemented by an appropriate combination of hardware and software. The software should be carefully implemented in order to maximize the effectiveness of the hardware. An operating system which supports relatively unlimited address or program space, is important. A program restart capability is important so that long execution times are not threatened by intermittent system failure. Powerful and efficient compilers will ensure that program execution will be as fast as possible.

Functional

The SAR/Landsat data merging system will have the capability of processing SAR and Landsat MSS images so that they are registered to each other and to a Universal Transverse Mercator (UTM) map projection. This is accomplished by first processing the Landsat data and then processing the SAR data to register with the corrected Landsat data. A user can then treat the corrected SAR and Landsat data in multispectral classification and other information extraction processes.

There are four major requirements for the SAR/Landsat data merging system. First is the need for data. Second is a system of inputting that data for use on a system prior to registration. Third is registration of the data. And lastly, the display of the data throughout the processes of input, registration, and final output states is required.

Data Sources - The sources of SAR data are proposed to be the Seasat satellite and aircraft. Landsat data may be acquired from any of three similar Landsat satellites launched between 1972 and 1978. SAR data are not always available immediately in digital format. In this case, the capabilities to form an image from SAR analog data as well as digitize the resulting image are necessary. Input may then be made to the reformatting software. Normally radar image formation will be accomplished by highly specialized equipment. The process is known as correlation of the radar data. The user is in general expected to furnish correlated digital SAR image data.

Data Input - Data input occurs in two steps. First there is the reformatting of an image to a common or standard computer format. At LARS this format is known as LARSYS format. Both Landsat and SAR data are transformed into this format.

The second step may be called data enhancement. Here the data is prepared for further processing. Enhancement may be broken down into cosmetic, radiometric and geometric processes. These processes overlap and complement each other. For instance, a cosmetic fix of a band line may have special radiometric considerations. Radiometric considerations are part of any magnification method applied to the data. As well, radiometric parameters must be considered in any geometric enhancement. An example is the transformation of the Landsat data to some specified grid with true north orientation. Pixel spacing and resolution are considered here. Algorithms must be available for manipulating these variables to aid high quality image formation, as well as for control point location and grid definition.

The specific Landsat processing functions that the system will have are:

1. Radiometric correction
2. Geometric correction

Radiometric correction is defined to be a striping reduction process that is performed on MSS data that has been radiometrically calibrated but not resampled. Geometric correction is defined to be the resampling of MSS data to a UTM map projection. The user will be able to specify both the spacing between pixels in the output UTM image and the orientation of the image (i.e., the direction of north relative to the image scan lines).

The SAR processing functions fall into the same two categories but different operations are performed. Specific radiometric correction processes that will be available for SAR are shading removal and spatial filtering to reduce speckle. A generalized polynomial distortion representation for geometric correction will be included.

Data Registration - The registration processing for SAR/Landsat imagery must satisfy several requirements. These include numerous image enhancements, both manual and automated control point location, an error tracking and/or modeling function and a registration process with various transformation algorithms for location matching as well as radiometric interpolation.

Image enhancements may be furnished from the data input phase. Among the enhancements needed are various filters to smooth data response from side to side as well as top to bottom of a data set, instantaneous field of view, principal components processing, and thresholded gradient. These will aid image interpretation and support further processing.

Locating control points in the images from the two imaging system may be accomplished manually or automatically. Manual control point location requires an excellent imaging capability which will be discussed further under the topic of data output products. Alternately automated correlation algorithms may be used. Several different methods may be attempted to achieve the desired accuracy.

Actual data registration will depend on the error tracking and modeling. The control points as well as systematic parameters of each scanner may be input to these error tracking or modeling programs. A transformation using various polynomials should be available as output. Data registration may follow this step.

The data registration program should give a registered data set in which the scanner images have a common grid. This grid may be a reference standard different from either of the input scanner data grids. Alternately, one of the two images grids may have been previously selected as the reference standard and the other scanner data subsequently registered to that. Further, the reference standard may be a map to which one or both images have been registered.

Data Output - The results of the merging system may be transmitted to the user in a significant way through the use of selectively chosen quality data output products. Each level of the data merging system may have specific requirements for the display of data. Permanent or semi-permanent images are needed to view data at various stages of the process. Locating control points is a function performed at the intermediate stage. Images may be displayed giving the before and after transformation comparison. Graphics may be used to show the final error.

Data output capabilities should include characters, to matrix gray levels, to false color photographs. Furthermore, the output will include inexpensive work copies, more expensive digital black and white photographs, and fine quality digital color maps. A temporary image displayed on a CRT may be useful at any step along the way.

Hardware

The SAR/Landsat merging system can be implemented on the IBM 370/148 computer at the Laboratory for Applications of Remote Sensing (LARS) at Purdue University. The only hardware requirements for the system are the tape drives needed for image data sets and sufficient virtual memory for the resampling program. The system will be designed to minimize the use of disk space by using tapes wherever it is reasonable to do so.

The resampling program will require an internal input buffer large enough to store the maximum number of input image lines that will be needed to create one output image line. This is a function of the rotational difference between the input and output images. Therefore, the memory requirements vary from image to image.

A scanning microdensitometer is needed for converting SAR data in film format to digital data. A variety of output devices can be used to produce images products. These include a line printer, CRT, high resolution paper product image output and high resolution film output. All but the film read/write hardware is resident at LARS but outside suppliers can be used to provide this service until a film writer becomes available. Resident service is desirable to ensure timely interaction with the user's data.

Software

The system will be written in Fortran IV and IBM370 Assembler Language. Generally, Fortran will be used wherever it is reasonable to do so. However, Assembler Language will be used for those programs that would be significantly more inefficient if written in Fortran. A highly efficient and dependable compiler as the IBM Fortran H is desirable to optimize the complex program functions.

The programs will use some local LARS programs to perform certain standard functions. For example, the program TAPOP will be used to perform image data set I/O.

User Considerations

Users of the SAR/Landsat data merging system have several choices to make when processing a Landsat data set. These user options are discussed next.

In some cases, a user will have a choice of using partially processed or fully corrected input MSS data. Since fully corrected data are already resampled to a standard map projection, further resampling may not be necessary. A common reason for further resampling would be the need for a different pixel spacing than the 57 meters (horizontal and vertical) of the fully corrected data set. Although resampling the data only once is clearly desirable, a recent study (Reference 7) has indicated that a second resampling will degrade the data only slightly and will not essentially change multispectral classification results.

When processing uncorrected MSS data and some partially processed MSS data, a merging system user must decide whether to use geodetic control points in determining the geometric transformation between the corrected and uncorrected spaces. If a systematic correction will provide satisfactory geometric accuracy to meet a user's needs and if no control-point library exists for a particular scene, then the expense of control-point location may be avoided by using systematic correction.

The system will have the capability to resample with nearest neighbor or cubic convolution. Although cubic convolution is widely regarded as a better algorithm than nearest neighbor, it is considerably more expensive to perform on a general purpose computer. Potential resampling algorithms for the merging system are summarized in Table 19. Computer expense for resampling is directly proportional to the number of pixels that are being processed. CPU estimates for the direct algorithms are extrapolations based on running times for an experimental prototype program on an IBM 370/168 Model 3 computer. Estimates for the hybrid-space algorithms are further extrapolations from the direct-program results.

Pixel spacing in the corrected Landsat data is a user option in the system. Uncorrected MSS pixels are spaced at about 57 meters horizontally and at about 79 meters vertically.

Table 19. ESTIMATED IBM 370/148 CPU TIME FOR RESAMPLING

Description of Resampling Algorithm	CPU Minutes per Million Output Pixels
Nearest Neighbor (Direct)	1.7
Cubic Convolution (Direct)	7.7
Cubic Convolution (Direct) with approximate high-frequency MSS correction	11.5
Cubic Convolution (Direct) with exact high-frequency MSS correction	15.0
Cubic Convolution (Hybrid space)	(3.9)
Cubic Convolution (Hybrid space) with exact high-frequency MSS correction	(5.0)

Fully corrected MSS data produced by the MDP will be spaced at 57 meters in both directions. The spacing in the images of Data Sets 2 and 3 was 25.4 meters. This particular spacing was chosen because the Seasat SAR data has a similar spacing. A second factor that a user must consider when choosing the pixel spacing is resampling cost. The number of output pixels that are created in the resampling step is inversely proportional to the pixel spacing.

Another user option for Landsat MSS processing is the orientation of the image. Orientation is the angle between north and an image scan line at some point in the image. The standard orientation is such that the spacecraft velocity vector is approximately perpendicular to the image scan lines. That is, input-space scan lines are nearly parallel to output-space scan lines. Orientation affects the resampling step, i.e., the memory required for the resampling program is directly proportional to the angle between the input space scan lines and the output-space scan lines.

Many of the user considerations for the SAR data will be similar to those for processing Landsat data. The user will have several options during the processing of his data, and these will be described next.

The first will be cosmetic in nature for bad data lines, i.e., a special case of radiometric correction. The primary method available will be repetition of the immediately preceding good data line. Other algorithms may be made available.

Second, radiometric corrections will be available in terms of shading corrections both along and across track as well as a low pass speckle removing filter. The user can elect to use none or all of these. At the present time, they are recommended as a package for the current aircraft SAR imagery.

Third, registration between the SAR and corrected Landsat data may be done via several polynomial functions or a parametric function. An evaluation of these functions will help the user determine whether a parametric, linear, biquadratic, or potentially higher order fit is required. Initially, the choices will be parametric, linear and biquadratic. Cubic and higher degree functions are planned for future phases of the project. Cost of registration is proportional to the complexity of the polynomial used.

Fourth, the resampling of SAR data to the corrected Landsat data pixel lattice is similar to the Landsat resampling. Essentially, use of the nearest neighbor interpolation will be considerably less expensive than the preferred cubic interpolation. The total expense is dominated by the resampling costs.

Fifth, images of SAR or Landsat may be produced in several ways as lineprinter, graylevel, printer plotter, CRT and filmwriter. When identifying control points, however, the highest quality images are required. Generally, filmwriter output is best. It is important to request such outputs very early in the process if special reformatting and outside vendors must be contacted and coordinated to produce the outputs.

Implementation at LARS

The Landsat/SAR data merging system will be implemented at LARS so that the user community may submit their corresponding individual Landsat and SAR images and obtain a set of registered multisensor digital images. Users will assist in the registration process especially during control point acquirement and data verification.

Implementation of this system (See Figure 78) is expected to occur in stages over several years. In the first of three stages, the most critical portion of the SAR/Landsat merging system will be implemented. Reformatting, geometric correction, and selected filtering of Landsat data is included along with SAR data reformatting as well as selected basic filtering. Limited imaging facilities present in this stage are to be augmented during later stages. Manual control point acquirement is to be utilized. Distortion correction functions available during the first stage will include systematic, affine, biquadratic and bicubic. The most important part of the system, registration, will be available to provide up to third order registration and resampling to include nearest neighbor and cubic interpolation. Phase I will provide a solid foundation; all essential requirements of the SAR/Landsat data merging system will be met.

The proposed system, diagrammed in Figure 78, is broken down by processing levels. These processing levels include reformatting, geometric correction, filtering, imaging, control point location, distortion evaluation, registration and resampling. These levels constitute the system elements to be implemented and documented by LARS. Some of these elements require extensive programming and documentation while other elements consist

APPROACH-MERGING SYSTEM DATA FLOW

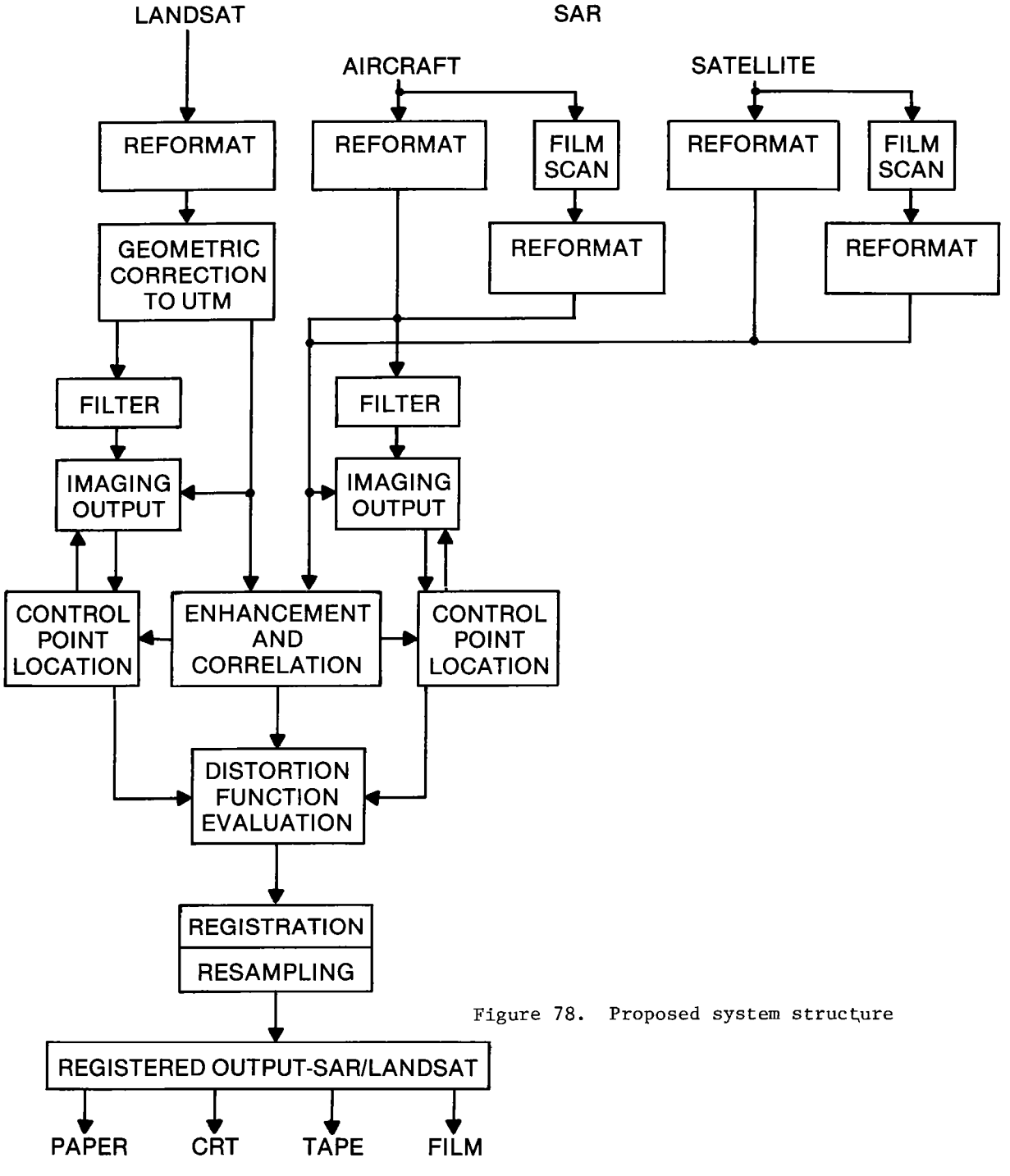


Figure 78. Proposed system structure

primarily of documentation. A film scanner/writer and a color video display are useful in the system but are not available at LARS, so external (to LARS) sources will be utilized until these items do become available. The following sections will discuss the system elements of Figure 78.

Input Data Reformatting

Reformatting of the Landsat data will be handled by the IBM software, but verification of this software will be necessary. SAR data reformatting programs exist and will need further refinement, standardization and documentation. The program refinement and standardization includes generalization and documentation. The program refinement and standardization includes generalization, for a variety of SAR data set inputs, addition of a free format control card reader, and internal program documentation. Program documentation oriented toward external users will be generated as input to a SAR/Landsat merging users guide. Phase II will implement Seasat and other satellite SAR data reformatting capabilities.

Landsat Geometric Correction

The next level of processing is geometric correction. Landsat data will be processed through IBM software to a UTM grid. Verification of the installation and operation of the software will be a task performed by IBM Federal Systems Division in cooperation with LARS. The Landsat MSS processing software which will be implemented at LARS will be defined in this section. Two options will be defined: Option 1 is a full Landsat MSS processing capability, while Option 2 is a minimal system that satisfies the geometric correction requirement. Option 2 is a subset of Option 1. Each option is described below.

1. Option 1 - Full Landsat MSS Processing Capability - The Landsat MSS software will consist of five programs:

   - Reformatting Program (see previous section)
   - Automatic Control-Point Location Program
   - Manual Control-Point Location Program
   - Geometric Transformation Program
   - Resampling Program

The inputs, functions, and output of each of these programs are described below.

There are three kinds of input MSS data that can be processed by the system:

- Uncorrected Data - This consists of four bands of MSS data in the X-format. The image data has been radiometrically corrected and line-length adjusted, but no other corrections have been applied.
- Partially Processed Data - This consists of four or five bands of MSS data in BSQ format. Only a radiometric correction has been applied to the image data. Ancillary data includes geometric transformation information.

- Fully Processed Data - This consists of four or five bands of MSS data in the BSQ format. Both radiometric correction and geometric correction have been applied to the data.

The SLDMS will be able to geometrically correct and perform a striping reduction on both uncorrected and partially processed data. It will also be able to resample fully processed data in order to change the pixel spacing.

As shown in Figure 79, there are three paths through the software. The first path involves the reformatting, automatic control-point location, geometric transformation, and resampling programs. This path would normally be used for an uncorrected MSS scene for which a corresponding control-point library exists. The result from this path is a fully corrected MSS scene with user-selected orientation and pixel spacing. This would be the most desirable path for processing uncorrected data. It can also be used to correct partially processed data. If no control-point library exists for an uncorrected MSS scene, there are two alternate paths through the system that can be used.

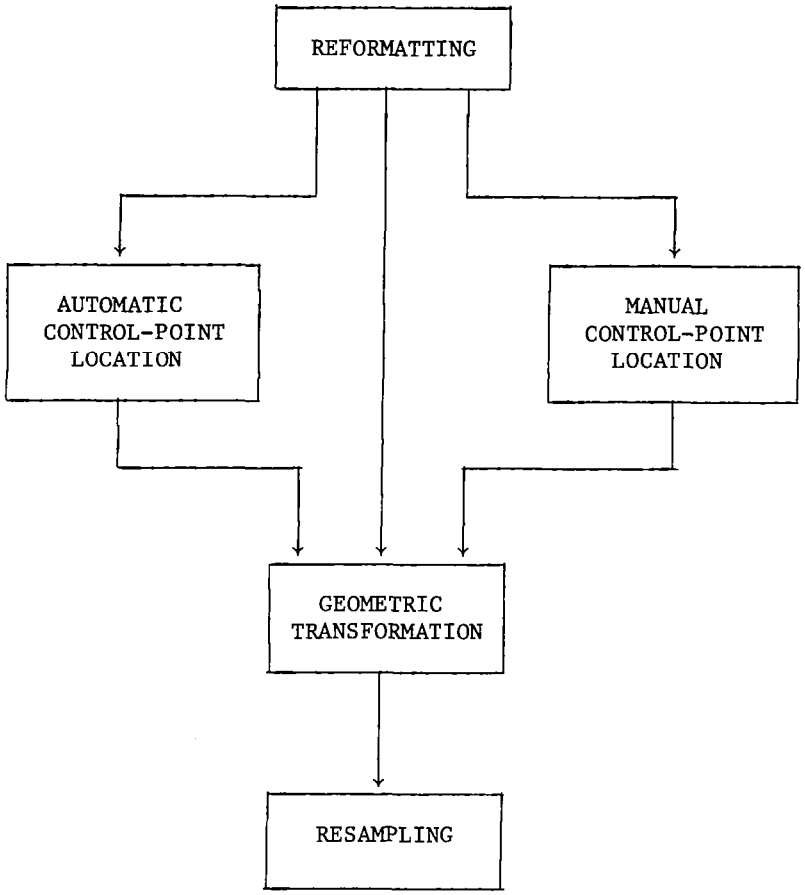


Figure 79. Option 1 software paths

The first and simplest alternate path involves only the reformatting, geometric transformation, and resampling programs. This path would normally be used for an uncorrected MSS scene for which no control-point library exists or for a partially processed MSS scene. The result from using this path on an uncorrected MSS scene would be a systematically corrected MSS scene with user-selected orientation and pixel spacing. From a partially processed scene, a fully corrected MSS scene that has been either systematically corrected or scene corrected (depending on the original ancillary data) can be produced with this path through the software.

The remaining path involves the reformatting, manual control-point location, geometric correction, and resampling programs. This path would normally be used for an uncorrected MSS scene for which no corresponding control-point library exists or for a fully processed MSS scene. The result from this path would be a fully corrected MSS scene. For a fully corrected input scene, only the reformatting and the resampling portion of the manual control-point location program are used. In this case, the result is a fully corrected scene with different pixel spacing.

2. Option 2 - Minimal Landsat MSS Processing Capability - The Landsat MSS software will consist of three programs:

- o Reformatting Program (see previous section)
- o Geometric Transformation Program
- o Resampling Program

The inputs, functions, and outputs of each of these programs are described below.

There are three kinds of input MSS data that can be processed by the SLDMS:

- o Uncorrected Data - This consists of four bands of MSS data in the X format. The image data has been radiometrically corrected and line-length adjusted, but no other corrections have been applied. Ancillary data includes no geometric transformation.
- o Partially Processed Data - This consists of four or five bands of MSS data in the BSQ format. The only correction applied to the image data is radiometric correction. Ancillary data includes geometric transformation information.
- o Fully Processed Data - This consists of four or five bands of MSS data in the BSQ format. Both radiometric correction and geometric correction have been applied to the data.

This option will allow the user to geometrically correct both uncorrected and partially processed data. It will also allow resampling of fully corrected data in order to change the pixel spacing.

The result for an uncorrected MSS scene would be a systematically corrected MSS scene with user-selected orientation and pixel spacing; from a partially processed scene, an MSS scene that has been either systematically corrected or scene corrected (depending on the original ancillary data) can be produced; from a fully corrected input scene, the

fully corrected scene with different pixel spacing.

The grid formed by the corrected Landsat will be used as the reference for all SAR data. Thus, the SAR data will be corrected to this grid during the data merging. Phase II will investigate Seasat geometric qualities. Phase III would implement results obtained from Phase II investigations of Seasat data.

Film data scanning and digitizing is expected to be a basic part of SAR data handling. Film scanning capability is not provided by the Phase I system implementation; however, this capability is provided by a number of organizations including NASA Wallops. A reformatting capability is included in this task for handling digital data from film scanners. Software will be implemented to provide a basic structure for reformatting any data format.

Image Filtering

Filtering of Landsat and SAR data is important to obtain the best possible imaging outputs as well as to bring out fully data analysis potential. Several image filtering methods were evaluated in the previous study and certain of these will be implemented. For SAR processing there are two which appear to be of significant value. One is an along and across track data normalization to reduce shading in the SAR imagery. The second is low pass filtering to partially remove speckle due to the nature of the SAR imaging process. Speckle is best reduced at the source by averaging several "looks" taken within the same resolution cell. If,however, imagery having speckle is on hand and the user wishes to process it, it may be desirable to spatially filter the data. These capabilities will be implemented. Other operations will be made available.

Imaging

Imaging is one of the most critical functions of the SAR/Landsat data merging system. Currently rudimentary outputs are available from paper, film and digital display. The paper outputs include lineprinter and gray level Varian printer/plotter. The lineprinter has low utilization while the gray level printer/plotter output has medium to high utility. The black and white digital display falls in the same category of medium to high utility. The available film recorder has medium utility because of its low resolution and small format. A more versatile color digital display would add measurably to the data merging system. Of high utility to the project would be in-house film scanning and film writing capability. Color film writing is highly desirable. Current external vendor capabilities provide good experimental turnaround, but will not meet the needs of an operational system. High quality imaging is critical to the control-point location process. Phase II would include procurement of a color digital display and digital film scanner and color film writer. Phase III would refine the implementation of these for better user interaction.

For Phase I, as part of the system document, procedures to use the available imaging devices will be described.

Control Point Location

Control point location will be a manual process with some automated assistance. Current imaging capabilities both in-house and out-of-house will be used fully to provide the quality of image needed to obtain accurate checkpoint locations. The high resolution film writing outputs have been exceptionally important to successful location. The electronic table digitizer may be used to measure the locations found for corresponding checkpoints. This process will be further refined and user documented during Phase I. Phase II will see slight revisions for use of new equipment to be procured during this phase. Further investigation will be made into automated cross correlation between SAR and Landsat data.

Distortion Evaluation

Image distortion evaluation will be carried out using least squares fits to a number of mathematical models. In this study parametric, affine, quadratic, cubic, fourth and fifth order models were evaluated against the aircraft SAR registration problem. Results indicate the parametric, affine and quadratic representations be implemented in the Phase I system. These functions will enable the user to remove distortion and register most aircraft SAR data sets to a geographic reference. Higher order functions can be implemented later should the need be demonstrated.

The functions to be implemented are for the aircraft SAR image data. They are similar to the Correction of Landsat imagery to a cartographic reference discussed in the section *"Landsat MSS Image to Map Registration."*

Distortions in Satellite SAR imagery have not been discussed; however, it is expected that the general polynomial representations which will be made available will enable subsets of satellite frames to be registered to a Landsat reference. Further study will be required to define all distortions in Satellite SAR imagery.

The distortion functions proposed for implementation here will be integrated into a user oriented system and documentation generated which will allow a technically trained person to easily utilize the system.

Registration and Resampling

Registration is the heart of the SAR Landsat data merging system. Parts of the software for this function exist but in geographically separate locations and in experimental states. During Phase I a new modular registration system is to be implemented. Internal program documentation will be brought up to a high level as will the user oriented external documentation. Streamlined buffering and location calculation software will assure more economical computer usage. A registration scheme using a high order polynomial with linear interpolation will be implemented, where high order initially will probably be up to fifth. The system will be designed so as to minimize actual data manipulation. During Phase II refinements will be added to include advanced resampling schemes and a higher order registration if required.

Resampling will be accomplished simultaneously with registration. Resampling schemes to be made available include nearest neighbor and cubic interpolation. These will be designed to be modularized parts of the registration software. Appropriate internal and user oriented external documentation will be produced.

SUMMARY OF RESULTS

The results of a SAR/Landsat MSS Data Merging System (SLDMS) study are discussed. This system study was conducted to determine the algorithms and procedures necessary to merge aircraft SAR and Landsat MSS imagery. A system plan was then developed from this system study whose results apply to a flat terrain. Next, the design of a SLDMS based on the system plan was pursued. The SLDMS was tailored to the processing facilities available at the Laboratory for Applications of Remote Sensing (LARS) of Purdue University, and so to an installation at LARS. The Landsat processing portion of the system was designed in more detail than the SAR processing portion because more background and expertise was available for Landsat MSS than for SAR data handling.

Three aircraft SAR images were registered to the corresponding Landsat MSS scenes, and these three sample data sets became the subject of an experimental investigation; several methods for the registration of these three data sets were studied and the results presented. The radar data was supplied by the U.S. Air Force from an X-band Goodyear AN/APQ-102A SAR. The radar data was delivered to Wallops as phase histories stored on film.

The registration of the Landsat image to a map and of the SAR image to the corrected Landsat image for the first data set was accomplished by a systematic approach. This data set was analyzed using LARSYS, and the information extracted from Data Set 1 was more than that obtained by using the sensors individually. Proceeding on this acquired knowledge, two more SAR/Landsat data sets were assembled.

Digital techniques were developed and used to create registered aircraft SAR and Landsat MSS digital images in a UTM projection for two more data sets. The registration algorithms investigated for Data Sets 2 and 3 included several polynomial functions. Edge enhancement to more easily identify features was also tested.

The common Landsat MSS frame for these scenes was resampled and geometrically corrected (using geodetic control points) to a 50.8 meter pixel spacing. Two subimages of this frame were then corrected to a 25.4 meter pixel spacing. The first subimage was of a ten mile strip of land extending from Salisbury, Maryland to Georgetown, Delaware. An aircraft SAR image was registered to this Landsat subimage (Data Set 2). The SAR image had adequate resolution but insufficient contrast for effective image registration, i.e., control points were difficult to locate. The speckle which was plainly evident in the SAR images of Data Sets 2 and 3 precluded the use of classifiers that use pixel-by-pixel methods of analysis. A biquadratic and a cubic function were used to perform the registration. The RMS residual errors of the least squares fit at the RCP locations of the biquadratic version were as large as 9.3 pixels. The least square fit of the cubic version had reasonable registration at the 14 RCP locations; however, some severe misregistration is evident at locations between the RCPs.

The second subimage was of an area which included Cambridge, Maryland. An aircraft SAR image was registered to this second Landsat subimage to give Data Set 3 in which there was sufficient resolution and contrast for image registration. The RMS residual error of the least squares fit at the sixteen RCP (reduced set) locations of the biquadratic version was 1.8 pixels.

CONCLUSIONS

This study has shown that digital methods provide the capability to accurately register synthetic aperture radar (SAR) imagery with Landsat multispectral scanner (MSS) imagery so that both overlay a map. The registration accuracy can be better than one Landsat MSS pixel. While the computer technology for this process exists, it is diffuse; that is, the expert at SAR image processing has little experience with MSS data. The registration system was not available at any single location. As the need for this multichannel data is perceived, the user will be driven into developing this capability for himself. There is, therefore, a need to implement this merging system at an accessible location.

The results from the investigations of the three SAR/Landsat MSS data sets indicate that the registration of SAR imagery with Landsat MSS imagery is both feasible, from a technical viewpoint, and useful, from an information-content viewpoint. The information extractable from a SAR/Landsat data set is greater than that obtainable from either sensor alone. Multisensor registration is likely to be a very desirable process in the future. It will enable users to make use of varied sources of information to solve their problems. This study demonstrates the viability of using SAR data in this way.

The investigations of the three aircraft SAR/Landsat MSS data sets clearly demonstrated that the greatest difficulty in registering aircraft SAR data to corrected Landsat MSS data is the location of control points. The difficulties occur in the following areas: selection of control points (features), proper distribution of control points, comparison of Landsat and SAR control points and error-free determination of control point locations. The selection of features that will serve as good control points is made difficult by the differences in the SAR and MSS imagery. Features that are usually high-quality control points in the MSS data (such as road intersections) were often unusable in the SAR imagery. Features of normally lower quality in MSS imagery such as agricultural field boundaries, had to be used. This problem is compounded by the fact that features often looked quite different in the data from the two sensors. For example, the shadowing of the SAR data did not occur in the MSS data. It is necessary that the set of control points be well-distributed throughout the image. In all cases investigated, there were geometric errors in any relatively large area that contained no control points. During the registration process, the primary use of reconstructed imagery is for locating ground control points. The image must have the quality of a photograph in order to identify features; once

identified the grid location of the feature must be easily obtained. These requirements conflict and several images may be needed, i.e., one image with small pixels for feature identification and one with large pixels for feature grid location. Data output products are important at each level of this data merging system. All of the problems normally associated with checking for errors in control point locations are present in the SAR/Landsat registration problem. Care must be taken to avoid obvious errors, e.g., all control points near the edges should be double checked since errors in the location of these points can easily go undetected. Errors in control point location caused the first two versions of Data Set 3 to have severe geometric errors. Most of these control point location errors were of features located near the border of the SAR image. When a set of error-free control points was used, the resulting corrected image had good registration with the corrected Landsat MSS image.

The experience from the investigations of the SAR and Landsat data sets indicate that SAR/Landsat data sets are unsuitable for automatic computer correlation of digital control point data. It is clear that the sensor responses cannot be compared directly due to the different response characteristics of the SAR and MSS instruments. An attempt to use an edge detection algorithm on the SAR imagery was not effective for locating control points due to the speckle and shadowing.

The design of the SAR/Landsat Data Merging System described in this report applies to SAR imagery which has been corrected for terrain (elevation) variation. An implementation of this system is feasible on state-of-the-art processing facilities. The modifications of this system necessary for its handling of SAR satellite imagery and aircraft SAR imagery which contains significant geometric distortion due to terrain variations, needs further study. Changes in elevation have produced little geometric distortion in the Landsat MSS imagery, however the aircraft SAR will show more effect.

The approach to registration should be selectable by the user, and the user should assist in the registration process, especially during the location of control points. Several registration algorithms were investigated. More than one should be available to the users. For the case of the registration of short segments of aircraft SAR to corrected Landsat MSS imagery, the systematic approach was found to be adequate; i.e., a control point RMS accuracy of better than one Landsat pixel was achieved. A systematic algorithm in which the coefficients in the transformation equations are related to physical variables such as skew and scaling, was found to be a basically simple approach. Future studies should take account of whatever ancillary information is available to improve the accuracy and reduce the cost of SAR registration. This approach is extensively exploited in the Landsat MDP correction system with resulting high accuracy and stability. Better SAR registration accuracy was achievable with higher order polynomials but at a cost of processing time and simplicity. However, the total expense is dominated by the resampling costs. The use of nearest neighbor rather than the much used cubic convolution will result in considerably less costly resampling.

The speckle inherent in the SAR instrument did not seriously interfere with the registration process unless the speckle was accompanied by poor contrast. However, the speckle did interfere with computer analysis, i.e., the speckle disturbs any signature analysis which is based on a pixel-by-pixel method. It follows that multiple pixel blocks will probably have to be analyzed as an aggregated unit. It also follows that a SAR to UTM reference registration RMS error in the one to two SAR pixel range should be acceptable. This level of accuracy has been shown to be achievable using a first or second order polynomial function for a SAR data set which is free from anomalous distortion.

REFERENCES

1. Landsat Data Users Handbook, Revised, NASA/GSFC, Greenbelt, Maryland.

2. Chu, N., McGillem, C. Anuta, P.: Analysis of the Effects of Interpolation and Enhancement of Landsat-1 Data for Classification and Area Estimation Accuracy, LARS Technical Report 110477, 1977.

3. "Study of Synthetic Aperture Radar (SAR) Imagery Characteristics," Goodyear Aerospace Corporation, NASA CR-141398, May 30, 1975.

4. Odenyo, V., Pettery, D.: Land-Use Mapping by Machine Processing of Landsat-1 Data. Photogrammetric Engineering and Remote Sensing, vol. 43, No. 4, April 1977, pp. 515-523.

5. Maurer, H., Clemens, P.: Assembly and Analysis of SAR/Landsat Data Sets. Proceedings of the Purdue/LARS Symposium on Machine Processing of Remotely Sensed Data, June 1977.

6. Swain, P., Pattern Recognition: A Basis for Remote Sensing Data Analysis. LARS Information Note 111572, Lafayette, Indiana.

7. Benner, R. Young, W.: Bi-resampled Data Study. Final Report Contract NAS5-23708, IBM Corporation, March 1977.

8. Manual of Remote Sensing, American Society of Photogrammetry, Falls Church, Virginia, 1975, Chapter 9, p. 458.

9. Graham, L. C. Peterson, R. K.: Synthetic Aperture Radar. Goodyear Aerospace Corporation GIB-9215A, June 1975.

10. Murphrey, S. W.: Final Report for Contract NAS6-2827, NASA CR-156847, September 1978.

APPENDIX A

SYNTHETIC APERTURE RADAR TUTORIAL MATERIAL

Basic Principles of Synthetic Aperture Radar Operation

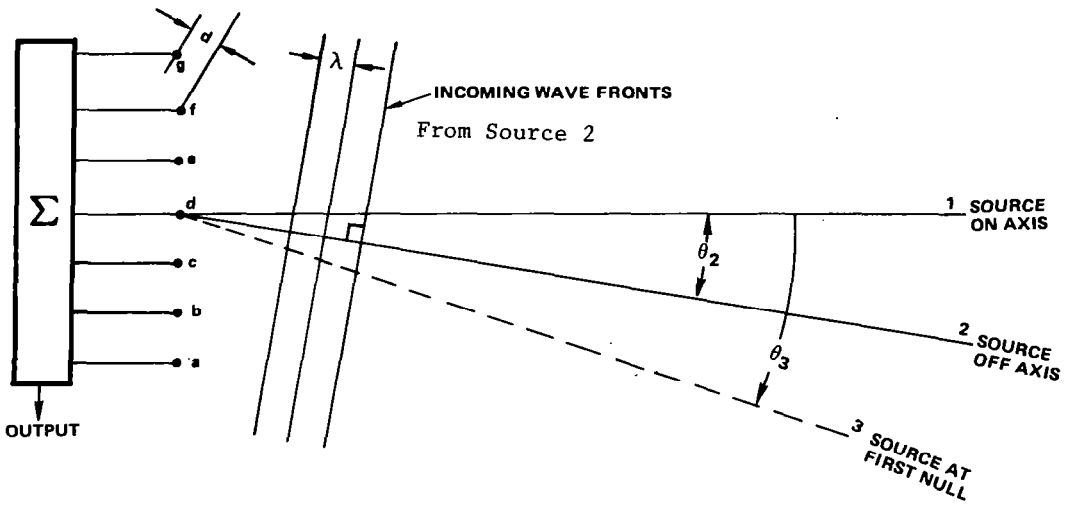
As discussed earlier, synthetic aperture radar (SAR) produces fine resolution images in the azimuth or along-track direction by storing signals received as the radar moves along a nominally straight line path and subsequently combining these stored signals to achieve the effect of a long aperture. The phenomenon involved is analogous to the formation of an image by a lens or of a receiving beam with a real aperture antenna. In such systems, signals received at different points across an aperture are vectorially summed, so that those arising from a point on axis are in phase and sum to a large value while those arising from some other direction are not in phase and sum to a lesser value. The relationship between angles of arrival and output is the antenna pattern.

The process is illustrated in Figure A1. Figure A1a shows (approximately plane) waves approaching an array or generalized aperture from a distant off-axis source. Sources are shown at three angles from the broadside axis; $\theta = \theta_1 = 0$, or broadside; $\theta = \theta_2$, or an arbitrary angle; and $\theta = \theta_3$, the first null angle.

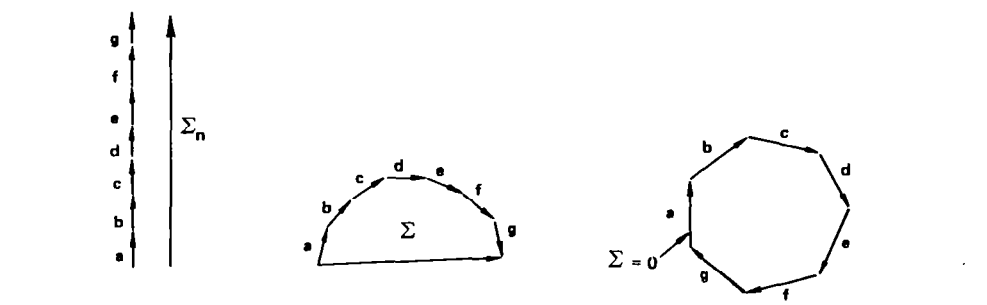
As shown in the figure, waves from various angles will arrive at the sample points of the aperture at different times because of the time delay required to travel the distance, d, so that they will not in general be in phase, except for those on axis arising from point 1. Figure A1b shows how the vectors sum from the three-point sources. The vectors from point 1 are in phase and sum to a maximum value, Σ_n. Those from point b sum to an arbitrary value less than Σ_n. In the special case of θ_3, the vectors sum to 0. At progressively larger angles, Σ goes through nulls and secondary maxima referred to as sidelobes. Figure A1c illustrates the total pattern.

If desired, the amplitude of the sidelobes may be decreased at the expense of some main lobe broadening by attenuating the signals away from the aperture center. This is referred to as amplitude weighting.

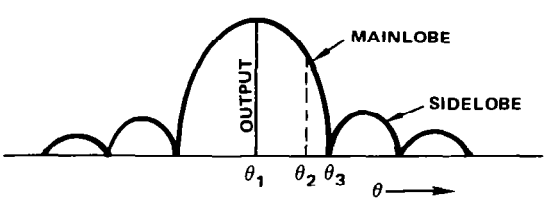
In synthetic aperture systems, essentially the same summation as in Figure A1 is performed, except that the signals are summed over time instead of instantaneously being summed over distance i.e. over the aperture. Thus, the signals must be stored as they are received until the required synthetic aperture has been generated. Although summation may begin as soon as the first signals are received, the image formation is not complete until all signals from the aperture have been received and summed. The general process is illustrated in Figure A2. As shown in the figure, a coherent source of microwave radiation is produced by momentarily connecting the amplified output of a stable local oscillator (STALO) to an antenna carried by a vehicle along a flightpath.



A1a Waves and Array Elements



A1b Array Output



A1c Output as a function of θ

Figure A1. Antenna pattern formation

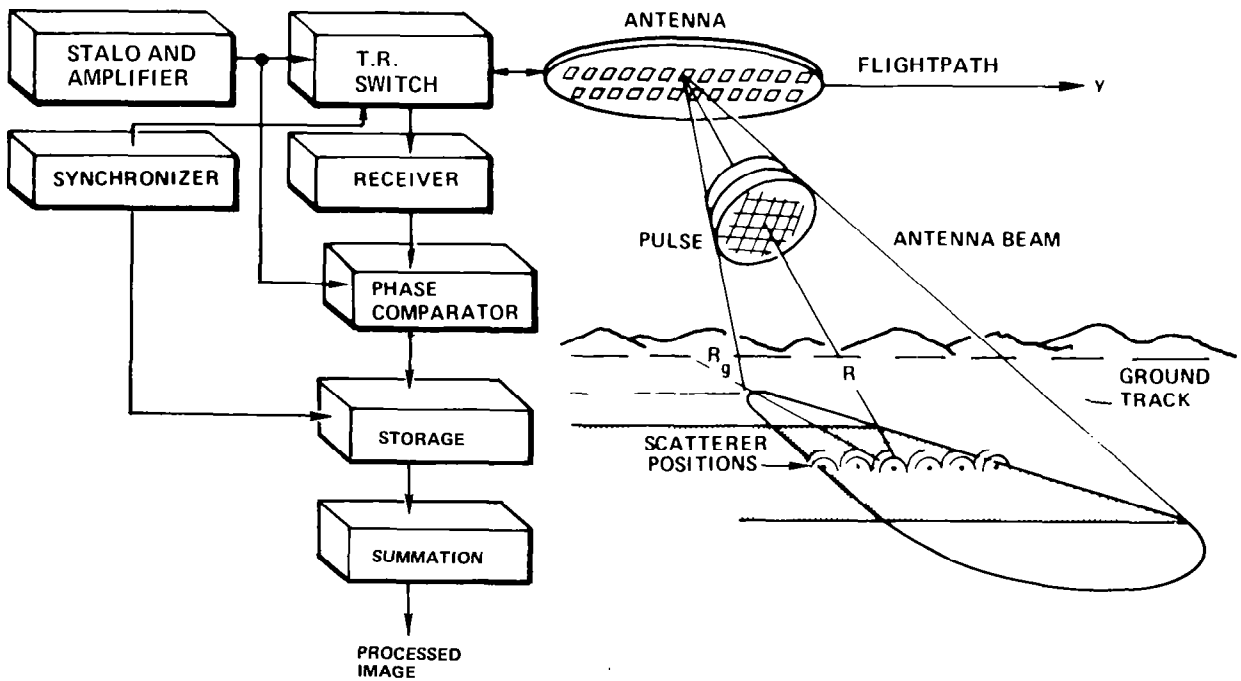


Figure A2. General synthetic aperture radar operation

The resulting pulses of energy are scattered by terrain elements; some of the energy returns to the antenna where it is amplified and compared in a phase detector with a signal derived from the STALO. Under the timing action of the synchronizer, those signals which arise from scatterers at the same range and thus arrive with the same time delay from successive pulses are stored together and subsequently summed. The processed image, in general, is data which relates one dependent variable, scattering strength, as a function of two independent geometric variables, usually distance from the flightpath and distance along track.

Storage and summation can be accomplished by any of a number of analog, digital, or hybrid processes. Two types of processing for synthetic aperture image formation will be discussed here. The first is coherent optical processing in which the signals are stored on photographic film and processed in a coherent optical analog processor. In the second, the signals are either inserted directly into a digital signal processor which contains sufficient storage in digital memory elements to satisfy the storage requirement-and arithmetically summed, or the signals are first stored on magnetic tape and subsequently read out into a processor of the same type.

Optical Signal Recording and Processing

A typical coherent optical storage and processing system is shown in Figure A3. In this system, the phase-detected signal is used to modulate a line scanning recorder. The imaged spot is swept as a function of time during reception of scattered energy. Thus, transmittance of the subsequently developed recorded strip from a single transmitted pulse is a function of scatterer reflectivities and the relative phase between the received signal and the phase comparator reference.

Returns from scatterers at other distances, or ranges, are recorded at different distances from the edge of the film. Returns from scatterers at different positions along track are recorded at different positions along the film, and may overlap signals recorded from other scatterers at the same range.

The phase relationship between transmitted and received signals for a particular target changes slightly from pulse to pulse because of the slight change in range to a scatterer as it passes through the beam. The range, and thus the phase difference, decreases from the maximum value occurring when the target first enters the beam to a minimum at broadside, and then increases again to approximately the same maximum value at the trailing edge of the beam. The rate of change of phase is the Doppler frequency. Figure A4 graphically shows the relationship between scatterer position in the antenna beam, range (phase), frequency, and phase-detected signal amplitude for successive pulses, which is recorded as transmittance along the signal film.

As shown in Figure A4 and the inset in Figure A3, the (approximately parabolic) phase change as the scatterer moves through the beam causes the phase-detected output to vary as the received signal goes in and out of phase with the transmitted signal, rapidly at

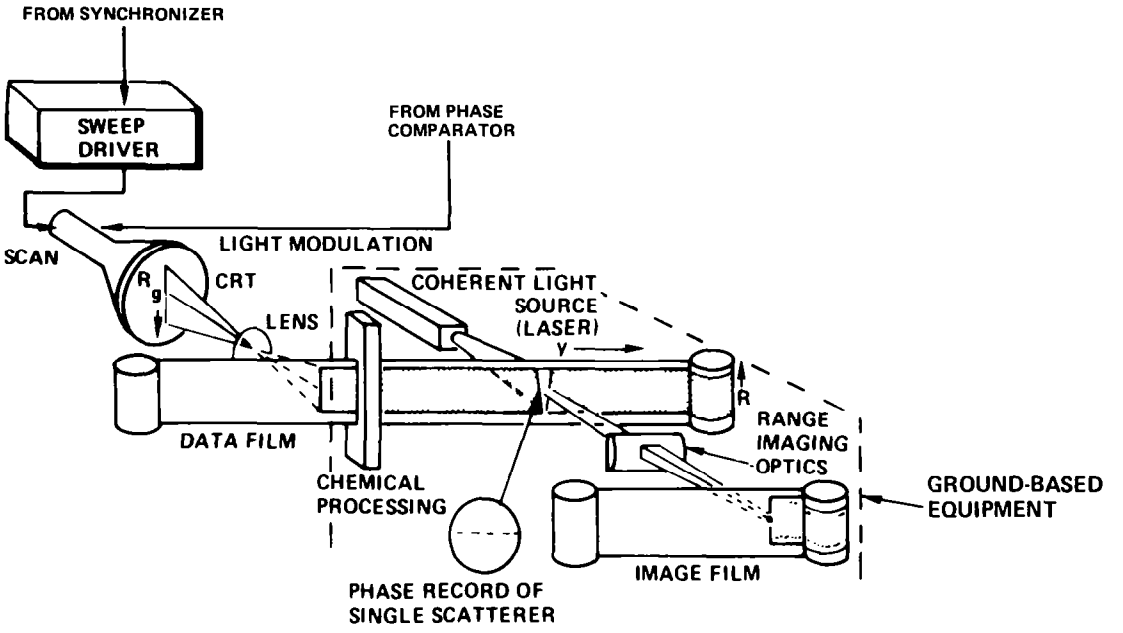


Figure A3. Synthetic aperture radar optical processing

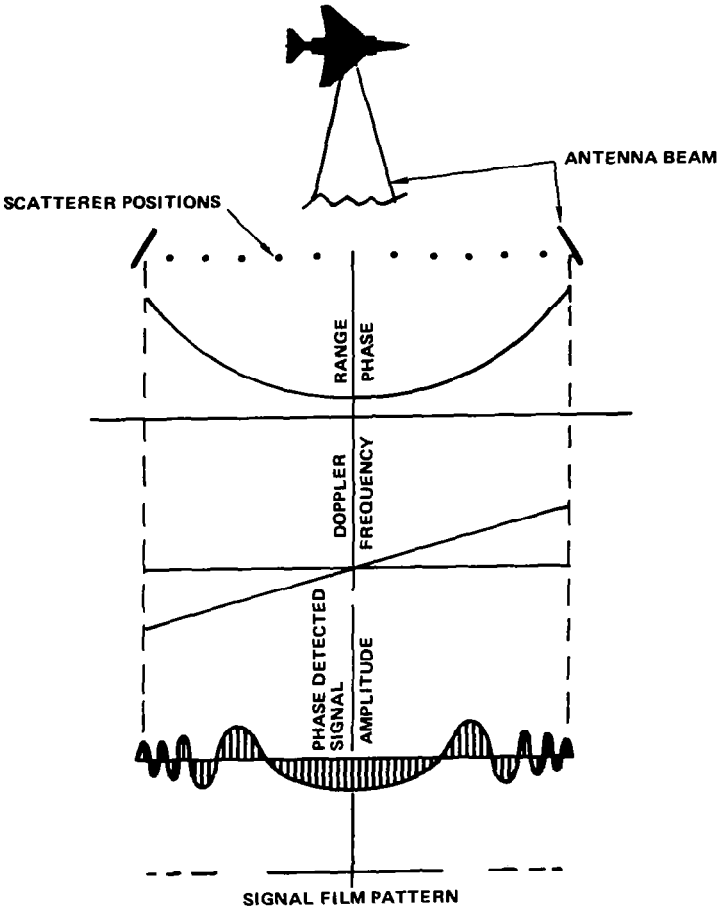


Figure A4. Phase detected and recorded signals

the edges of the beam and slowly near broadside. The resulting pattern of light and dark areas on the signal film acts as a section of a zone plate in the coherent optical processor which, along with anamorphic range imaging optics, focuses light from the coherent light source into a point on the image film. Patterns, or "phase histories," from other scatterers which lie at different ranges or along-track positions are simultaneously recorded at the positions--only a fraction of the available phase history is normally illuminated at one time, the length used corresponding to the required synthetic aperture length.

Not all of the light impinging on the signal film is diffracted into the spot representing the scatterer. Because the signal film transmittance cannot be negative, a large portion of the light passes through and would contribute to noise in the final image. Furthermore, note that the signal from the phase detector is a scalar which is only the magnitude of the vector defining the phase and amplitude of the received signal. In a given pulse or short series of pulses, therefore, there is not information as to whether the Doppler frequency is increasing or decreasing; i.e., whether it is in the front or rear half of the beam. Assuming that some particular portion of the phase history were selected for processing, some portion of an adjacent phase history would have the same Doppler frequency and thus contribute noise to the final image.

Both of these problems are eliminated by offsetting the phase detector reference by an amount equaling the Doppler frequency at the edge of the antenna beam, so that the recorded Doppler frequency is zero at the edge of the beam and twice the actual frequency at the other edge. The resulting signal film pattern (phase history) is illustrated in Figure A5. Compare this to the signal film pattern in Figure A4, and see Figure A12.

When the offset phase history is illuminated in the correlator, three wave fronts are produced--one undiffracted, one converging, and one diverging from a virtual image of the scatterer. Because the three have different apparent source positions, the desired one may be selected. This is illustrated in Figure A6.

Amplitude weighting is achieved by the coherent optical channel design. It is roughly equivalent to shaping the illumination of the coherent beam in Figure A6, but is actually accomplished in a different part of the processor.

Pulse compression is usually employed in synthetic aperture systems. In this case, the system range resolution is that corresponding to the bandwidth of the dispersed (transmitted)pulse rather than to the time duration of the pulse. Pulses are either dispersed before transmission by passing a narrow pulse through a network which exhibits a linear phase shift versus frequency, or are generated as dispersed pulses by sweeping an oscillator over the desired band of frequencies. On reception, the dispersed pulses may be compressed by passing them through a network having a phase-frequency characteristic inverse to that used on transmission. Alternatively, the received swept pulses may be recorded on the optical processor data film and compressed optically simultaneously with synthetic aperture formation.

Figure A5. Offset phase history.

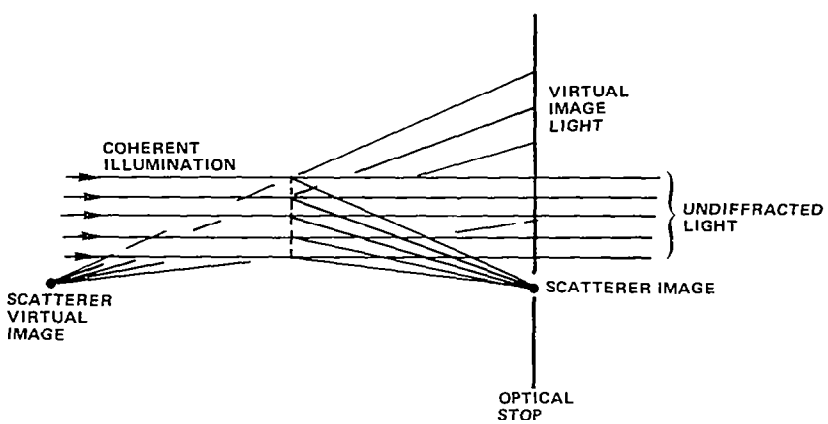


Figure A6. Elimination of undiffracted light and virtual image

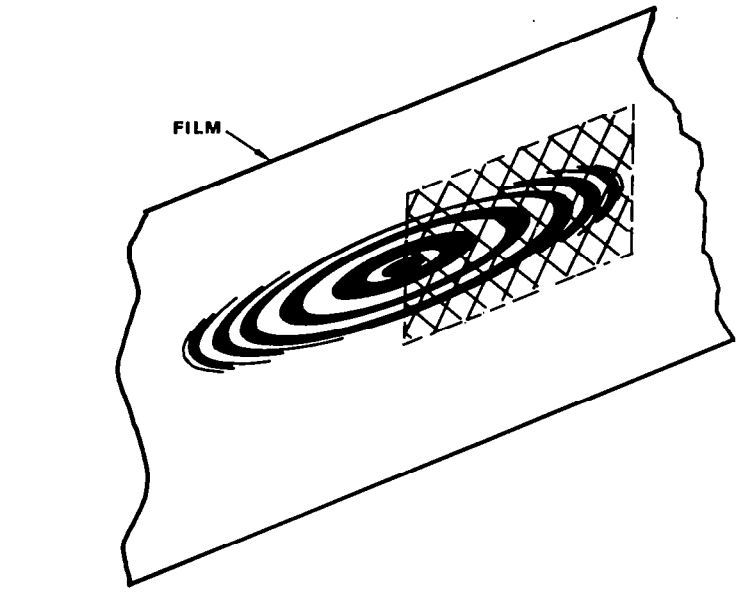
In this case, the recorded return from a single scatterer for one pulse is a phase history like that of Figure A4, except that it extends across the film instead of along the film. Successive pulses vary in phase as before, so that the complete recorded phase history occupies an areas on the film rather than a line. As will be derived later, the alignment of range phase histories is such as to cause the lines of equal phase to be hyperbolas or ellipses, depending on whether the frequency is swept up or down when the pulse is expanded. The two cases are illustrated in Figure A7a and A7b. Use of the offset scheme described earlier results in elimination of one-half of the pattern, as shown in A7a, and halving the spacing between rings at the high-frequency end. Alternatively, the phase detector reference may be offset by the IF bandwidth, in which case the top or bottom half of Figure A7a or Figure A7b would be used. When pulse compression is accomplished optically, weighting in the range direction, equivalent to that described previously for azimuth sidelobe control, is also accomplished in the range direction for range sidelobe control.

Digital Processing

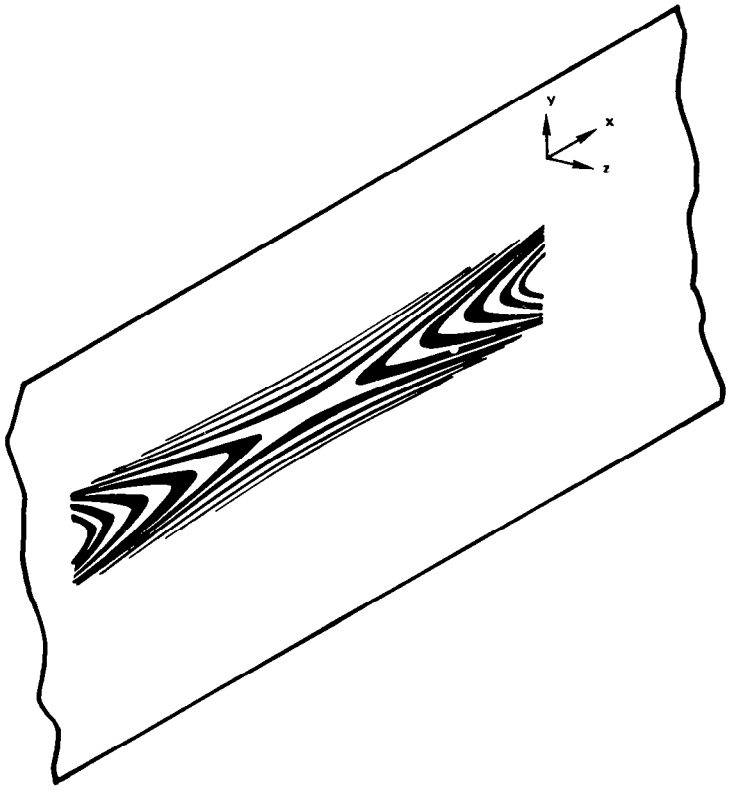
When the raw radar signals are to be processed digitally, the basic radar operation is the same as that described earlier. One important difference is that the various digital arithmetic algorithms do not suffer the analog limitation caused by the fact that film transmittance cannot be negative. The second problem which is solved in the optical system by offsetting the phase detector reference, that of eliminating the image of fore/aft ambiguity, may be solved in another way so that an offset is not required. This lowers the maximum Doppler (or alternatively, radar) frequency that must be recorded by a factor of two, at the expense of adding a second recording/processing channel.

The second channel is added as a quadrature channel. It is identical to the first receiving channel, except that the phase detector reference is phase shifted by $\pi/2$ radians. These two signals then form the components of the vector signal that has both phase and amplitude. One of the two receiving channels is therefore, arbitrarily designated I for "in-phase" and the other Q for "quadrature." In principle, each channel may be treated independently, forming a separate synthetic aperture and resulting image. Because of the arbitrary nature of the phase the image of an individual scatterer from a single channel may have an intensity anywhere from that corresponding to the scatterer strength to the negative of that, including zero. This effect, sometimes called in-phase/quadrature scintillation, is another manifestation of the inability to characterize the vector signals with a single value. The magnitude of the image vector is obtained by forming the square root of the sum of the squares of the individual I and Q channel output signals.

A typical digital signal processor block diagram is shown in Figure A8.



a



b

Figure A7. Film recording of Doppler phase history with range chirp

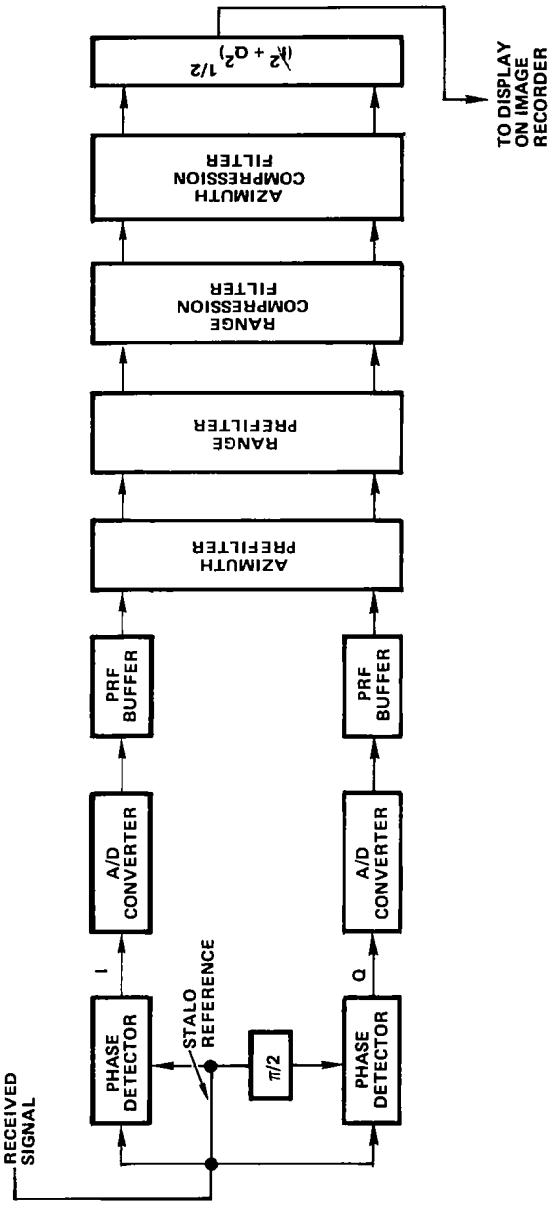


Figure A8. Typical digital signal processor

Digital Processing Operations and Algorithms

The individual processing steps in Figure A8 may be used to discuss the various processing operations and algorithms in general. The first step, of course, is to convert the analog signals to digital words. Thus, for each range resolvable element of the returns from each transmitted pulse, a word, usually four or five bits long, is generated to define the corresponding signal amplitude.

The bit stream from the analog-to-digital (A/D) converter exists only during the times that returns from the area of interest are being received. This normally amounts to 20 to 50 percent of the total interpulse period. The PRF buffer accepts the data bursts and outputs at a slower rate, reducing the data rates and bandwidths required for further processing or recording.

To avoid aliasing (ambiguities), the radar pulse repetition frequency must be more than twice the beam edge Doppler frequency. However, if not all of the synthetic aperture is required for subsequent processing to achieve performance parameters, the higher Doppler frequency data need not be used. This is removed from the data in the azimuth prefilter by conventional digital, low-pass filtering techniques. At the same time, aperture weighting for sidelobe control is applied. Weighting for range sidelobe control is accomplished in the range prefilter.

Several algorithms are available and used for performing pulse compression and synthetic aperture formation. A convenient and efficient algorithm for both when system performance parameters are not too stringent is convolution. When using this algorithm, the range-gating function is performed by passing each bit of the signal through a series of shift registers connected in series, as shown in Figure A9. Consider each parallel group of shift registers as a single one holding parallel words. Each of these is long enough to hold all of the detected returns from a single pulse. After all of the returns from the first pulse are into the first shift register, they begin to flow into the second and the returns from the second pulse start into the first shift register. After n pulses, sufficient for a synthetic aperture, the total phase history for any individual scatterer at a given range appears simultaneously at the interconnection between shift registers, and if attention is fixed on the values in the various shift registers at a given range, the phase history will of course travel through the entire shift register set. Now, if the swept pulse waveform to be compressed or the azimuth phase history expected for a scatterer at a given range, called the reference function, is multiplied by the signal sample for that range and the results added together over the expanded pulse length of the synthetic aperture, a maximum value will be obtained when the traveling phase history is aligned with the reference function. Lesser values will be obtained for other alignments, and the output as a function of alignment will be analogous to Figure A1c. Figure A1c and its implementation in an optical processor is usually thought of as a Fourier or Fresnel integral transform. The process in Figure A9

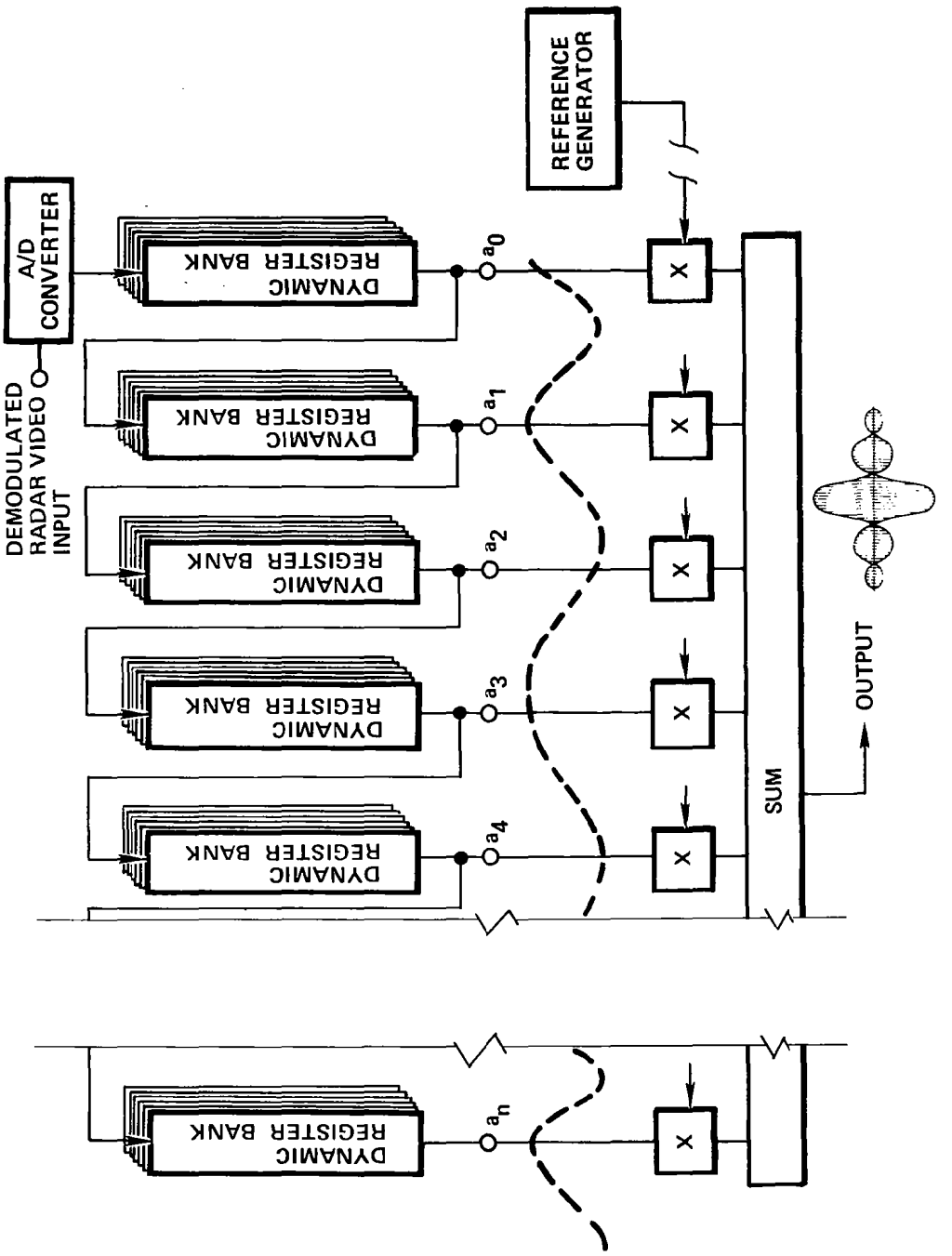


Figure A9. Radar processor using digital shift register storage

is usually thought of as cross correlation or convolution. Note that only the I channel is illustrated and that the reference function for the Q channel must be different (90-degree phase change). Both the range compression filter and the azimuth compression filter, which perform the functions of pulse compression and synthetic aperture formation, respectively, may be instrumented with the convolution algorithm.

Note, in passing, that range gating can also be accomplished by storage and readout of signals in a random access memory (RAM) and pulse compression, and synthetic aperture formation can also be performed by other techniques such as fast Fourier transforms (FFTs).

Derivation of Basic SAR Relations

Azimuth Resolution

An expression for the azimuth resolution of a SAR can be developed by considering the motion of the system through space on a platform as shown in Figure A10.

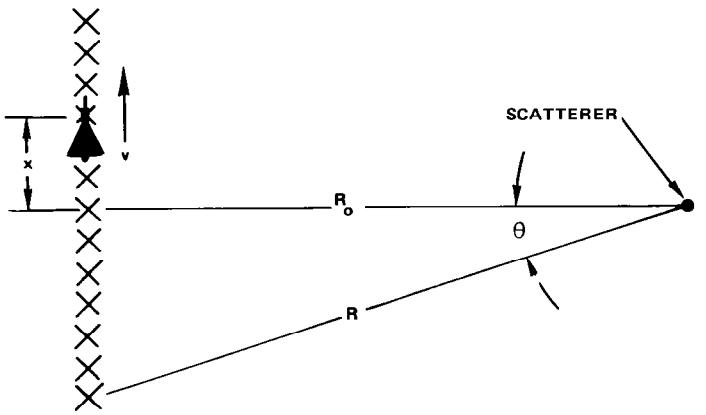


Figure A10. Synthetic aperture generation

Assuming that the radar transmits and receives energy from each of the positions along the flightpath, as shown in Figure A10, the phase shift of each point is

$$\phi = \frac{2\pi(2R)}{\lambda} = \frac{4\pi R}{\lambda} \qquad \text{(A1)}$$

and the range R as a function of distance along the track is

$$R = \sqrt{R_o^2 + X^2}$$

$$= R_o\left(1 + \frac{X^2}{R_o^2}\right)^{\frac{1}{2}} \approx R_o + \frac{X^2}{2R_o} \quad , R_o >> X, \tag{A2}$$

so that the phase is

$$\phi = \frac{4\pi}{\lambda}\left(R_o + \frac{X^2}{2R_o}\right) \tag{A3}$$

$$= \phi_o + \frac{2\pi X^2}{\lambda R_o}$$

This constantly changing phase represents a frequency shift of the returned signal relative to the transmitted signal, the Doppler frequency shift, given by

$$f_d = \frac{1}{2\pi}\frac{d\phi}{dt}$$

$$= \frac{2}{\lambda}\frac{X}{(X^2 + R_o^2)^{\frac{1}{2}}}\frac{dX}{dt}$$

$$= \frac{2V}{\lambda}\sin\theta \tag{A4}$$

This result is consistent with classical Doppler equations for the sum of the moving source and moving observer Doppler shift when the relative or vector velocity of the vehicle with respect to the scatterer is $V\sin\theta$. Also, the phase shift in the previous equation, except for the constant phase term, is exactly twice the function describing the phase shift of a focused linear array antenna. If the signals received from the scatterer can be stored, retaining both amplitude and phase, and then summed as in a focused array, a synthetic beam will be formed with beamwidth

$$\beta_s = \frac{\lambda}{2L_s} \tag{A5}$$

The length of the synthetic array generated by the moving aircraft is simply the arc extended by the beamwidth of the physical antenna, L_p, at a given range, R_o, or

$$L_s = R_o\beta_p = \frac{R_o\lambda}{L_p} \tag{A6}$$

The linear resolution of the synthetic aperture, W_s, is approximately

$$W_s = R_o \beta_s = \frac{R_o \lambda}{2(R_o \lambda / L_p)} = \frac{L_p}{2} \tag{A7}$$

as illustrated in Figure A11.

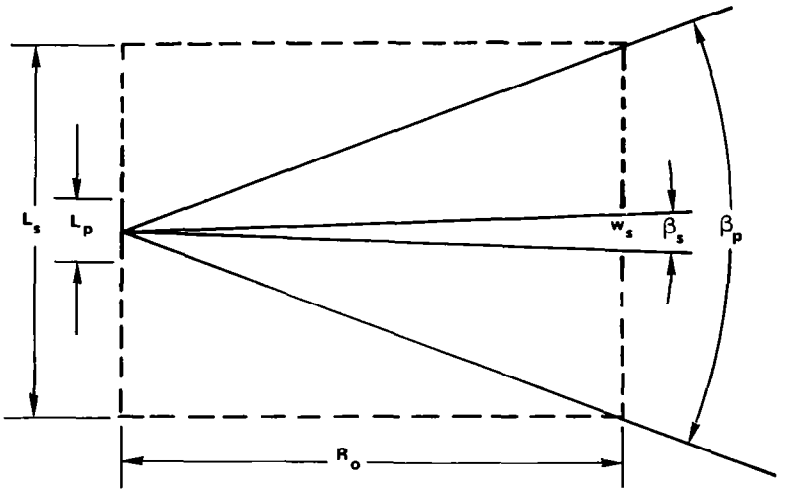


Figure A11. Synthetic aperture resolution.

Thus, limiting or best possible linear resolution of a synthetic aperture radar is taken as half the length of the physical antenna carried on the platform used to generate the synthetic array and is independent of range.

Unfocused Versus Fully Focused Synthetic Aperture Radar

A fully focused SAR must be able to keep track of the phase of the return from each pulse as well as to keep the synthetic array focused on the desired range to compensate for the change in range as the radar travels past a scatterer. The maximum phase shift, referenced to the shift when the target is normal to the flight path, occurs when the target just enters the beamwidth of the real antenna, or,

$$\Delta\phi_{MAX} = \frac{2\pi X^2}{\lambda R_o}\Bigg|_{x = L_{S/2}} = \frac{2\pi L_S^{\,2}}{\lambda 4 R_o} \quad . \tag{A8}$$

If this phase shift is less than $\pi/2$, that is

$$\frac{2\pi L_S^{\,2}}{\lambda 4 R_o} < \frac{\pi}{2}$$

$$\text{then} \qquad L_S < (\lambda R_o)^{1/2} \quad . \tag{A9}$$

It is not necessary to focus the synthetic array if inequality A9 is obeyed (Reference 7). Satisfying this inequality amounts to processing only the central portion of figure A4. The unfocused resolution is found from equations (A5), (A7), and (A9) to be

$$W_{S\ \text{unfocused}} > \frac{1}{2}\,(R_o \lambda)^{1/2} \quad . \tag{A10}$$

The fully focused resolution is $W_S = \frac{L_P}{2}$.

An unfocused radar system is considerably simpler and less expensive than a fully focused version. Processing can be accomplished by low-pass filtering the Doppler-shifted returns.

Rangewalk

The variation in range, as the platform travels past a scatterer, that produces the Doppler phase shift also varies the range position of the reflected pulse. If the range variation is an appreciable fraction of a range resolution bin, the return will appear at different range bins as the platform flies by. This is often referred to as rangewalk. Unless compensated for it can degrade azimuth resolution, range resolution, and system dynamic range. Rangewalk is generally a problem only in high-resolution systems and can be compensated for during processing.

Ambiguity Constraints

Range Measurement Ambiguity - In order to avoid the backscatter from successive transmitted pulses returning to the radar at the same time, the interpulse period, T, must be long enough such that

$$T \geq \frac{2\Delta R}{c} \quad , \tag{A11}$$

where ΔR is the range interval illuminated and c is the speed of light.

The pulse repetition frequency, PRF, is then limited by

$$PRF = \frac{1}{T} \leq \frac{c}{2\Delta R} \quad . \tag{A12}$$

Azimuth or Doppler Frequency Measurement Ambiguity - The Doppler frequency shift from Equation (A4) must be sampled by the PRF at a rate at least twice the highest Doppler frequency. The highest frequencies occur when

$$\sin\theta \approx \theta = \pm \frac{\beta_p}{2} \quad ,$$

and since

$$\beta_p = \frac{\lambda}{L_p} \quad ,$$

then

$$\theta = \frac{\lambda}{2L_p} \quad \text{at } f_{d\ max} \quad ,$$

so that from Equation (A4),

$$\frac{1}{T} = PRF \geq 2\left(\frac{2V}{\lambda}\right)\left(\frac{\lambda}{2L_p}\right)$$

$$= \geq \frac{2V}{L_p} \quad . \tag{A13}$$

The radar must be pulsed at least twice per antenna length travelled.

Combining equations (A12) and (A13) yields the ambiguity constraint for the PRF, which applies to two channel (I&Q) systems, as usually instrumented in digital processors.

$$\frac{c}{2\Delta R} \geq PRF \geq \frac{2V}{L_p} \quad . \tag{A14}$$

Spectrum Folding, Image Ambiguities - The time and frequency representations of the received Doppler frequency shifted signals from the geometry of Figure A1 are given in Figure A12. As shown, the signals as received from the scatterer at the moving platform

are first slightly higher in frequency than the transmitter, then the same, then slightly lower. The relative phase decreases to zero then increases to its original value. The video output from the synchronous demodulator contains frequencies which decrease to zero and then increase again. The difference between negative and positive Doppler shifts is lost as the Doppler spectrum is folded. To eliminate this, the phase detector reference frequency may be offset by an amount equal to the maximum Doppler frequency expected. The result is shown, is that the maximum Doppler frequency with offset is twice that in the previous case, so the sampling theorem requires that

$$PRF \geq 4f_{d\ max}$$

and equation (A13) becomes

$$\frac{c}{2\Delta R} \geq PRF \geq \frac{4V}{L_p} \quad \text{(azimuth offset)}. \tag{A15}$$

which is the form normally applicable to systems using optical processing.

In order to prevent spectrum folding with azimuth offset, the PRF must be doubled and the maximum slant range decreases by half. Or, by doubling the antenna length, the PRF and slant range can remain unchanged and the best possible azimuth resolution is degraded by a factor of two.

The offset can also be applied to the range information. Range offset, however, requires twice the video bandwidth.

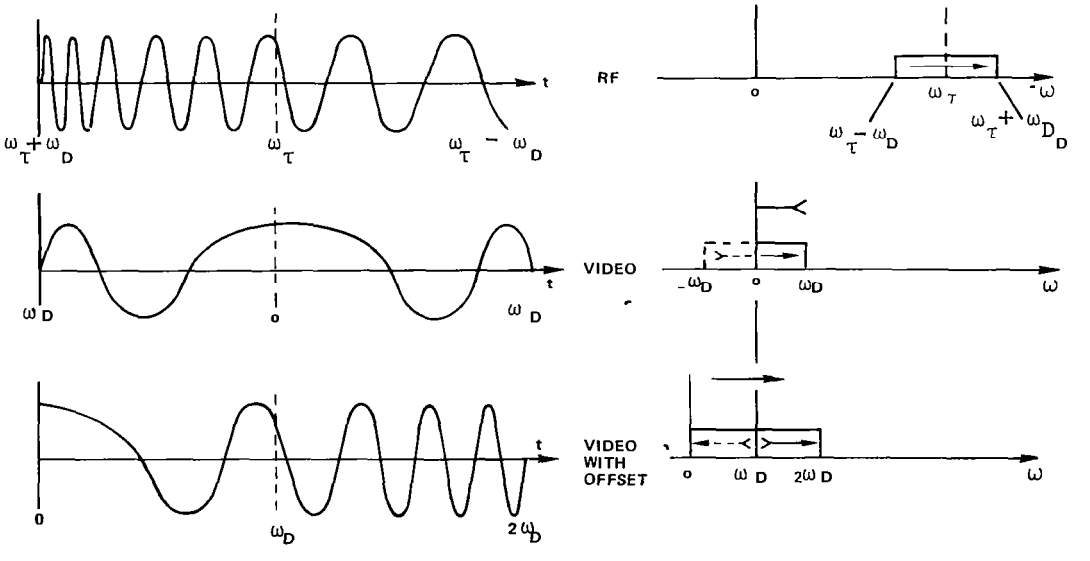


Figure A12. Spectrum folding, image ambiguities

Radar Equation for Synthetic Aperture Radar - The radar equation (Reference 8) for synthetic aperture radar relates the signal-to-noise figure for the received information to the radar and image parameters.

$$\frac{S}{N_c} = \frac{0.44\lambda(PRF)}{\beta_p V} \frac{P_t G^2(\theta)\lambda^2 \beta_p c\, \tau\, \sigma_o \sec\theta}{(KTB)(NF)2(4\pi)^3 R^3} \tag{A16}$$

where

$\frac{S}{N_c}$ = Signal-to-noise ratio for coherently processed backscatter

$G(\theta)$ = antenna pattern

θ = grazing (or depression) angle

λ = wavelength

P_t = transmitted power

β_p = antenna azimuth beamwidth

τ = pulse length

R = slant range

V = velocity of platform - ground speed

σ_o = radar cross-section of target

K = Boltzmann's constant

T = absolute temperature

B = system bandwidth

NF = system noise figure

PRF = pulse repetition frequency.

The term $0.44\lambda(PRF)/\beta_p V$ is due to the coherent integration of the multiple pulses that illuminate the scatterer.

The radar cross-section, σ_o, is the equivalent area of an isotropic scatter normalized by the actual illuminated area. The radar cross section is a function of frequency, polarization, and incidence angle, variables controlled by the radar, as well as many natural variables that include surface roughness, soil moisture, and vegetation cover. A curve of σ_o versus grazing angle for X-band is given in Figure A13.

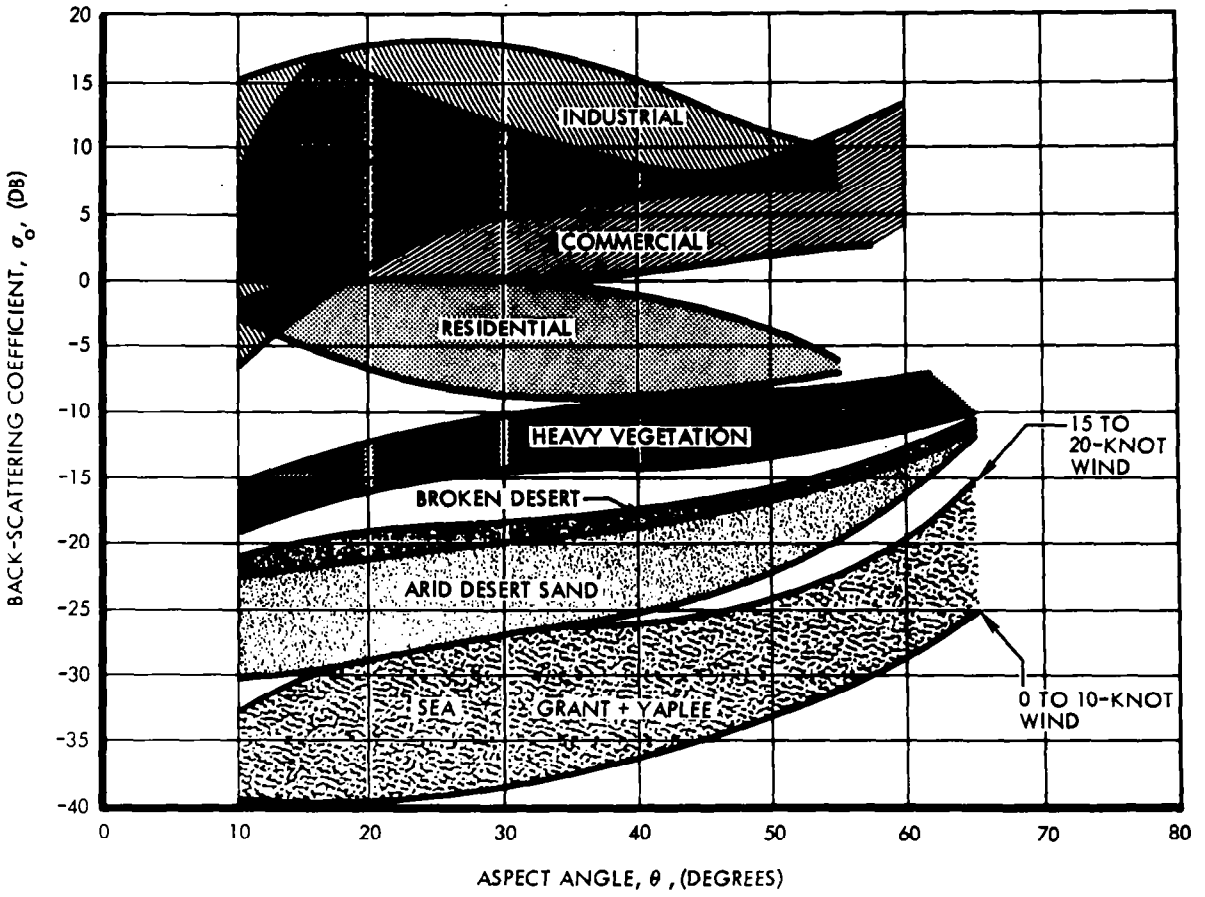


Figure A13. σ_o versus incidence angle, X-Band

Synthetic Aperture Radar Processors

Optical Processors

The first SAR processors were analog electronic devices using storage tubes for memory and one-pole RC or two-pole LC electronic filters as the processing elements. Resolutions attained, although a small fraction of the illuminated beamwidth, were 100 m or more. Later, optical data processors, more commonly referred to as correlators or correlator-processors, were conceived and demonstrated at the Willow Run Laboratory of the University of Michigan (now ERIM) and developed at Goodyear Aerospace Corporation for operational use. These transform holographic phase histories, which were recorded on airborne radar signal films, into radar images.

The recorded phase history of a single scatterer on the data film, with range chirp, would appear as in Figure A7. When illuminated with monochromatic coherent light, the range and azimuth patterns spatially modulate the light and, with the proper optics, focus to the point on the image plane corresponding to the target.

To determine the characteristics of the pattern on the data film, it is possible to combine the phase caused by aircraft motion (Equation (A3)) and that due to the frequency sweep of the expanded pulse. On any given return, the transmission delay from radar to scatterer and back can be thought of as controlling the starting phase of the expanded pulse, so that the two phases add:

$$\phi = \frac{\pi 4R^2}{c^2 \tau_e \tau_c} \pm \frac{2\pi X^2}{\lambda R_o} \quad , \tag{A17}$$

where

τ_e = pulse width of expanded pulse

τ_c = pulse width of compressed pulse.

Expressing this with data film coordinates,

$$\phi = \frac{4\pi k_1^2 S^2}{c^2 \tau_e \tau_e} \pm \frac{2\pi k_2^2 r^2}{\lambda R_o} \tag{A18}$$

where

k_1 = range scale factor on data film

k_2 = along-track scale factor on data film

S = range coordinate on data film

r = along-track coordinate on data film.

To characterize the pattern of the recorded data for a single scatterer, lines of equal recorded density can be traced; that is, equal phase. With ϕ constant, Equation (A18) is a family of ellipses for the positive sign, Figure A7a (frequency sweep and

Doppler frequency change in the same direction) and a family of hyperbolas for the negative sign, as in Figure A7b.

Fully automatic correlator-processors have been designed and manufactured for such synthetic aperture radar systems as the AN/APQ-102A and AN/UPD-4. Many of these devices, similar to the one shown in Figure A14, have self-contained film processing systems that enable the rapid development, fixing, and drying of data and image films. This type of unit has been used with a wideband data link to provide real-rate, near real-time processing and viewing of radar imagery. Film transport and chemical processing requirements necessitate approximately five minutes delay from radar illumination of an area to viewing the same area on the attached viewing table.

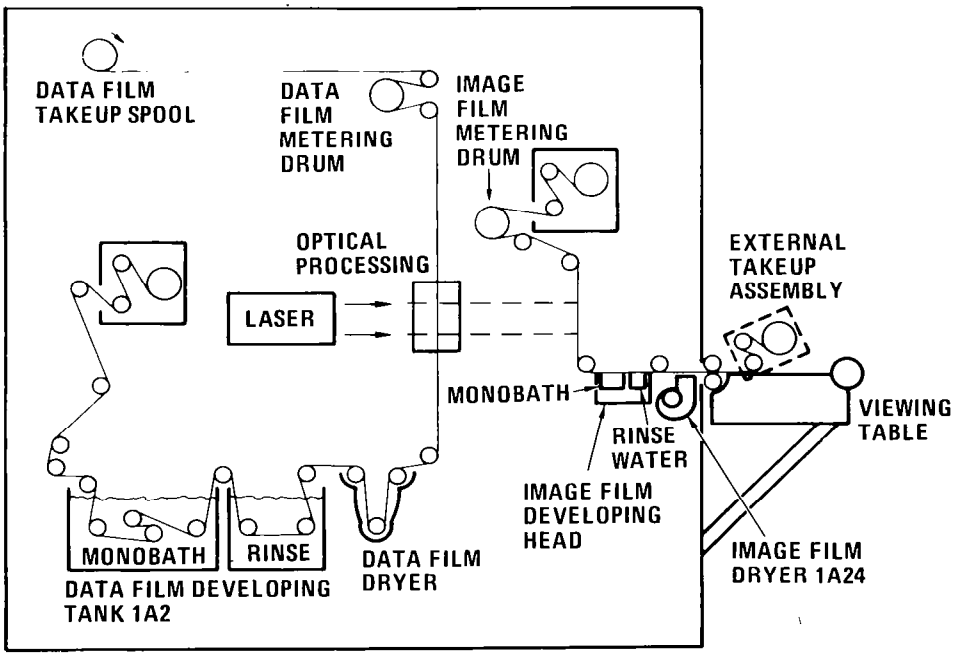


Figure A14. Correlator-processor functional diagram

APPENDIX B

IMAGE RECONSTRUCTION

Alternatives

Remote sensing data is often available in numerical, rather than image form. The numerical form lends itself nicely to computer aided analysis; however images of data still maintain an important role during data preprocessing, data analysis, and the presentation of results. During digital image registration the reconstruction of images from the numerical data is a key factor and can affect the accuracy of the registration.

Uses of Imagery

Initially, during the registration process, reconstructed images prove a valuable tool in assessing data quality. These images may be used to determine types of enhancements or corrections that need to be applied to the data. If in viewing the imagery, the data appears noisy one might wish to evaluate different filtering techniques. If bad data lines are evident in the imagery, the imagery may be used to determine which lines are bad. Some form of correction may then readily be applied to the data set. Banding or striping may be apparent in the reconstructed image, if so a striping removal algorithm may be applied to the data. In viewing the imagery, look angle or sun angle effects may be evaluated to determine the necessity of applying across track correction software. Along track the gain may vary, and the imagery can be used to assess the degree to which this occurs so an appropriate correction algorithm may be applied. Geometric distortions may also be evident in the image which might mandate the use of geometric correction software. Once any enhancements or correction algorithms are applied to the data set, reconstructed images of the enhanced data may be used in visually evaluating those enhancements or corrections applied.

In locating ground control points the imagery may be used to manually choose these points or at least to inform an automatic correlation algorithm of a starting area. During the manual process, a high quality image is needed in order to distinguish ground features to be used as control points. It is also necessary to identify individual pixels, whether directly or by use of a table digitizer.

After the data sets have been digitally registered, reconstructed imagery may be used to visually evaluate the registration. This may be done by producing a split-screen image (see figure B1) of the data sets where ground features such as roads may be used to roughly gauge the registration accuracy. Superimposed images of the registered data sets (see figure B2) may also prove valuable in evaluating the registration.

The data analyst may use the reconstructed images of the data sets to choose areas of interest. He may also use the imagery to pick training fields or samples to train the classification algorithm used, or to pick test fields or samples to evaluate the classification results.

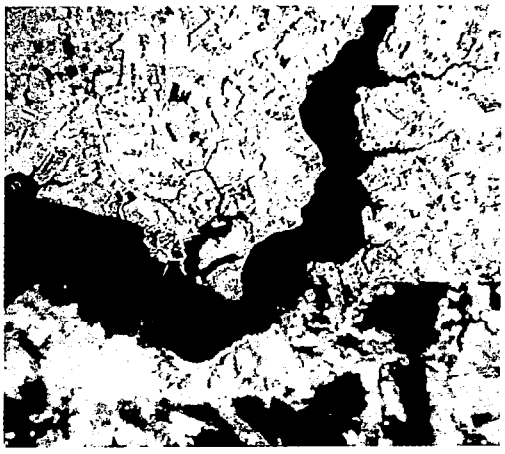


Figure B1. Split-screen gray scale image of registered data set upper portion SAR data, lower portion is Landsat spectral data

Figure B2. Gray scale image of SAR data superimposed on Landsat false color image using registered data set, produced on IBM's filmwriter

Devices

In generating a reconstructed image there are several devices from which to choose, each different advantages and disadvantages varying with the use which is going to be made of the image. The devices available may be grouped into four types which are normally used for image reconstruction: line printers, plotters, digital displays or CRT's, and film-writers.

A typical line printer used for image reconstruction is the IBM 1403 printer. In using a line printer to construct an image, typically characters of varying sizes and densities are used to represent different gray shades. A blank may be used at one extreme of the spectral response and a character such as M at the other extreme. One method sometimes used, in order to extend the range of the gray shades available, is to use overprinting to get darker shades of gray. When used as an image reconstruction device each character printed usually represents a pixel, thus on a standard line printer the produced image would contain ten pixels per inch horizontally and eight pixels per inch vertically. Figures B3 and B4 illustrate line printer images.

The major advantages for using a line printer as an imaging device are its high availability and low cost. Also point location is readily available at the pixel level which is very desirable during the process of locating control points.

Among the disadvantages is the poor feature recognition; at such a magnification some ground features are no longer easily identifiable. Also characters are not a visually pleasing method for viewing gray levels. There is the restriction of not being able to represent a solid black shade. When the data set is large the cumbersomeness of the printer image limits its usefulness.

The major use of the printer images is in identifying the pixels making up ground features whether for use as ground control points during registration or for use as training or test fields during data analysis. This use may be greatly aided by using a conventional image with small pixels in conjunction with the printer image. The line printer may also be used to create a classification map when it is desired to know how individual pixels were classified.

An example of a plotter used for image registration is the Varian Statos 4200 (see figure B5). The Varian electrostatic printer/plotter creates gray scale images by using dot matrix patterns to represent the different gray levels. The Varian Statos 4200 has 200 dots per inch, thus if a five by five dot matrix is used there would be 40 patterns or pixels per inch both horizontally and vertically. Example images are shown in Figures B6 and B7.

The Varian plotter has, among its advantages, the capability of varying the resolution of the output device by selecting different sizes of dot matrices with which to represent the data samples; thus image data with either an equal aspect or equal resolution along the X and Y directions may be imaged easily. The images are relatively inexpensive to generate so they may be used as working copies during the

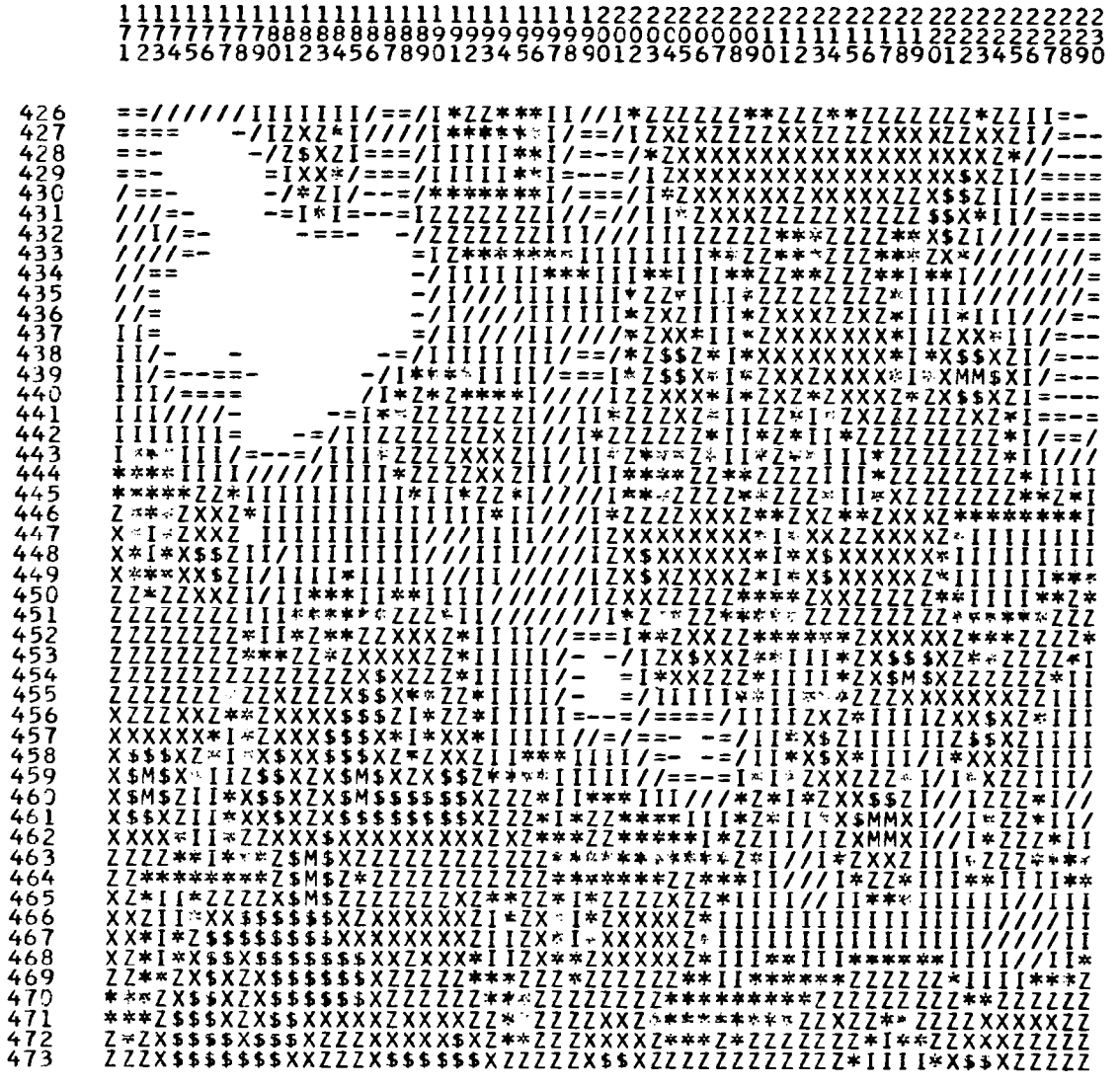


Figure B3. Gray scale image of Landsat spectral band 5, produced on line printer

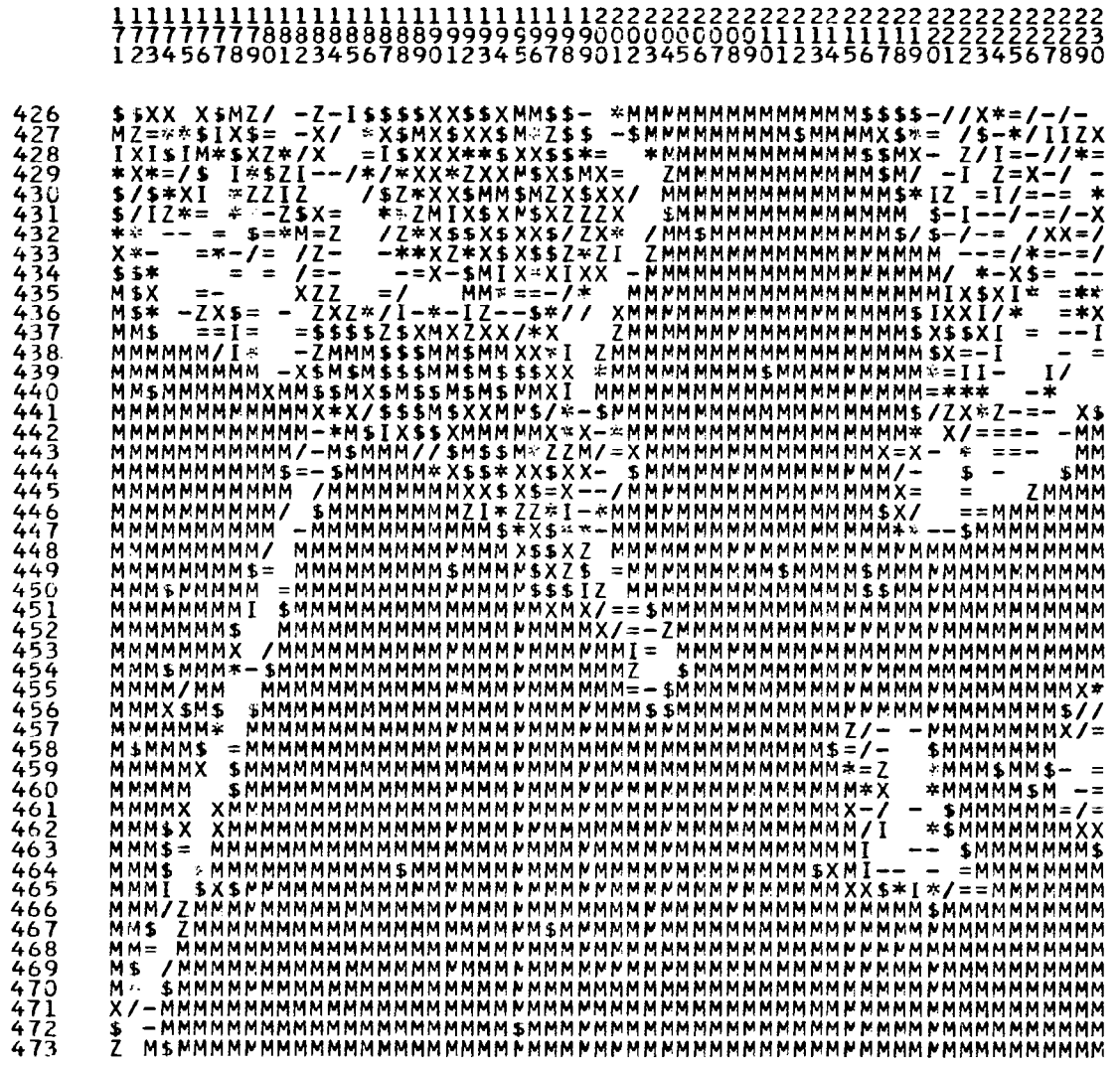


Figure B4. Gray scale image of SAR data, produced on a line printer

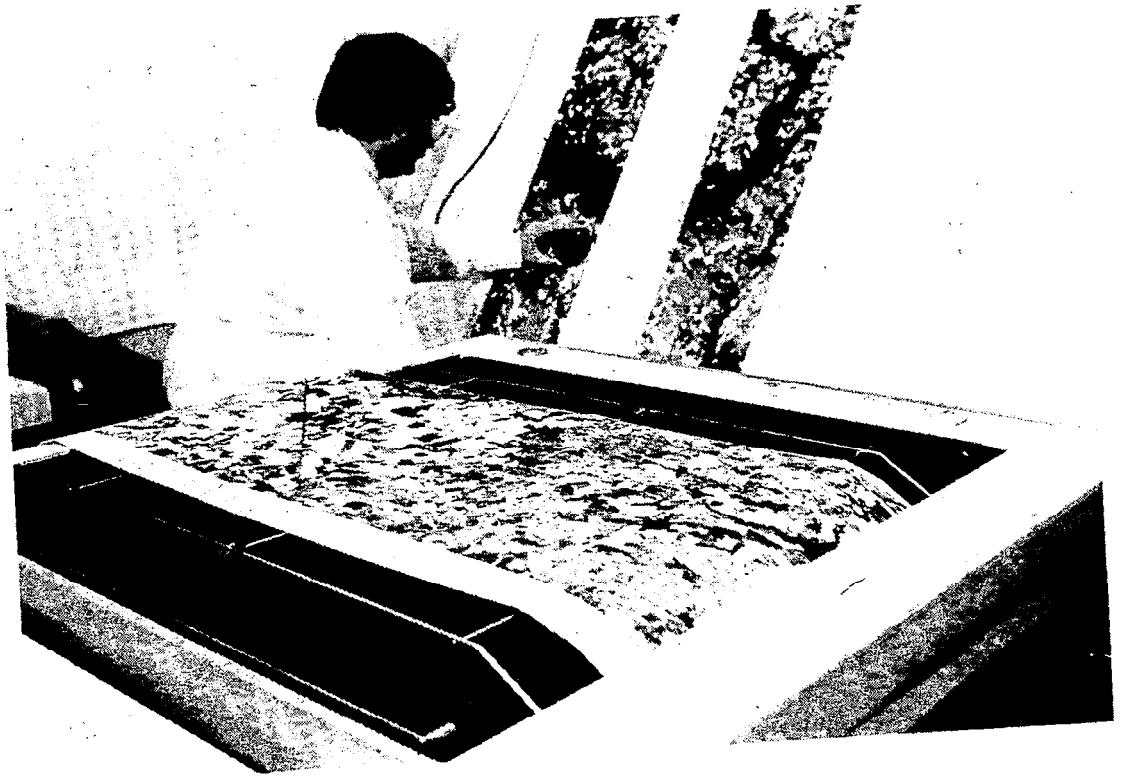


Figure B5. Varian statos 4200 electrostatic printer/plotter.

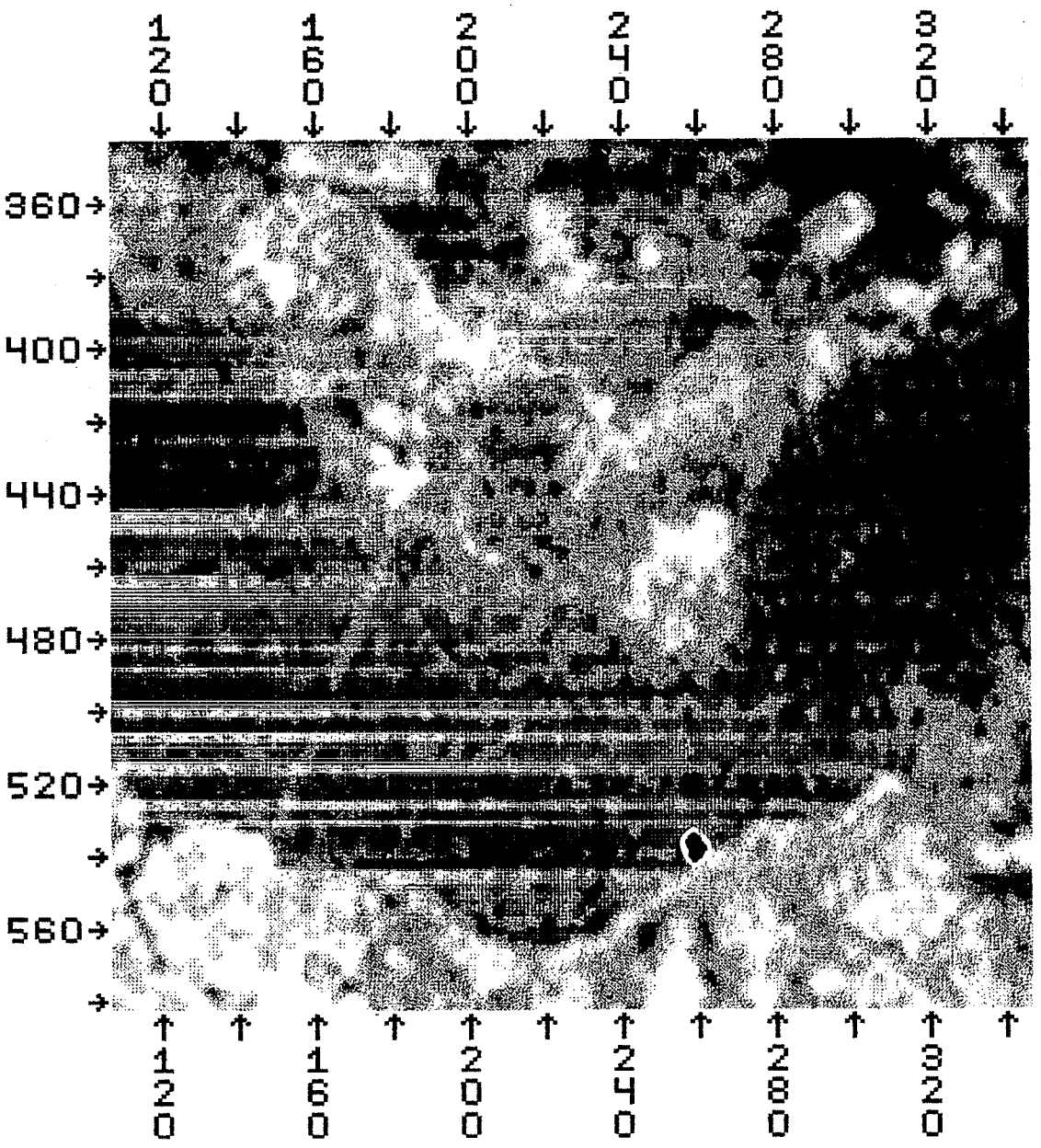


Figure B6. Gray scale image of Landsat spectral band 5, produced on a Varian plotter

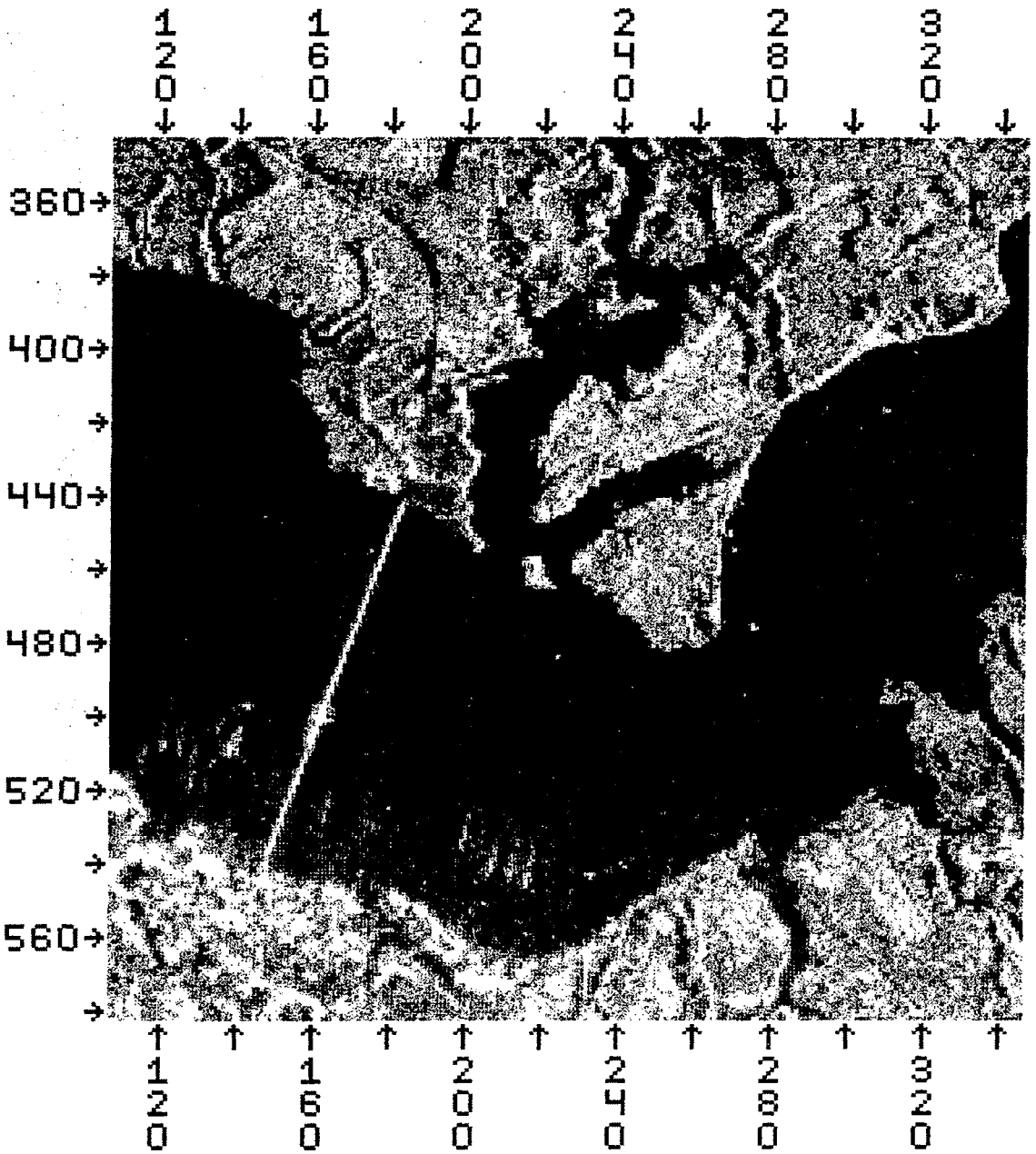


Figure B7. Gray scale image of SAR data, produced on a Varian plotter.

registration or analysis procedure. Also the image produced is of a quality suitable for reports or as a final output product. The entire area of interest may normally be viewed without undue difficulty, which adds convenience for the analyst whether registering the data or performing data analysis. When used in conjunction with a table digitizer, ground control points may be accurately located within the data set or ground features may be accurately delineated by the data analyst.

If a high resolution image is desired the number of gray levels available becomes limited. For example, if an image is desired with 100 pixels per inch, then two by two dot matrix patterns must be used. Thus the number of distinct gray levels, represented by different dot patterns, is greatly reduced. Also since the gray levels are represented by dot matrix patterns, occasionally some rastering occurs in large areas assigned to the same gray level. It is not well suited for visually assessing registration accuracy since imaging is limited to single channel output.

The Varian plotter images have a sufficient resolution so they may be used to assess data quality. They also may be used in accurately locating ground features. Since the Varian plotter images are economical enough to be used as working copies and the quality is usually sufficient for use as final output products they prove very versatile in their applications.

Another image reconstruction device is the digital display, an example of which is the IBM 4507 (see figure B8). The IBM 4507 consists of a pair of high resolution black and white T.V. monitors. The first is for viewing and manual interaction, while the second (a slave unit) is for photographic purposes. The interactive monitor has a 12 inch by 16 inch viewing area which can display 576 lines and 768 columns. Each pixel may be represented by one of 16 gray levels. Also associated with the digital display is a light pen and keyboard which aid in the interactive use of the display. Using the slave monitor one may photograph data which appears on the display. Photographs obtained from the monitor are shown in Figure B9 and B10.

The major advantage of using the digital display is that it may be used interactively, thus coordinates are available immediately. A fairly large amount of data may be viewed as a whole and interactive enlargements may be made of areas of interest. The gray levels are distinct and smooth in nature.

The major disadvantage of the digital display is cost. Reevaluation of an area requires reconstructing the image an additional time. The only permanent output available is from the photocopy unit which, since it is a photograph of a video display, may contain some distortions.

The digital display images have a variety of uses. The resolution is of such a nature as to aid in evaluating data quality. The display may be used to interactively locate ground control points or training and test fields. Since the display has a split

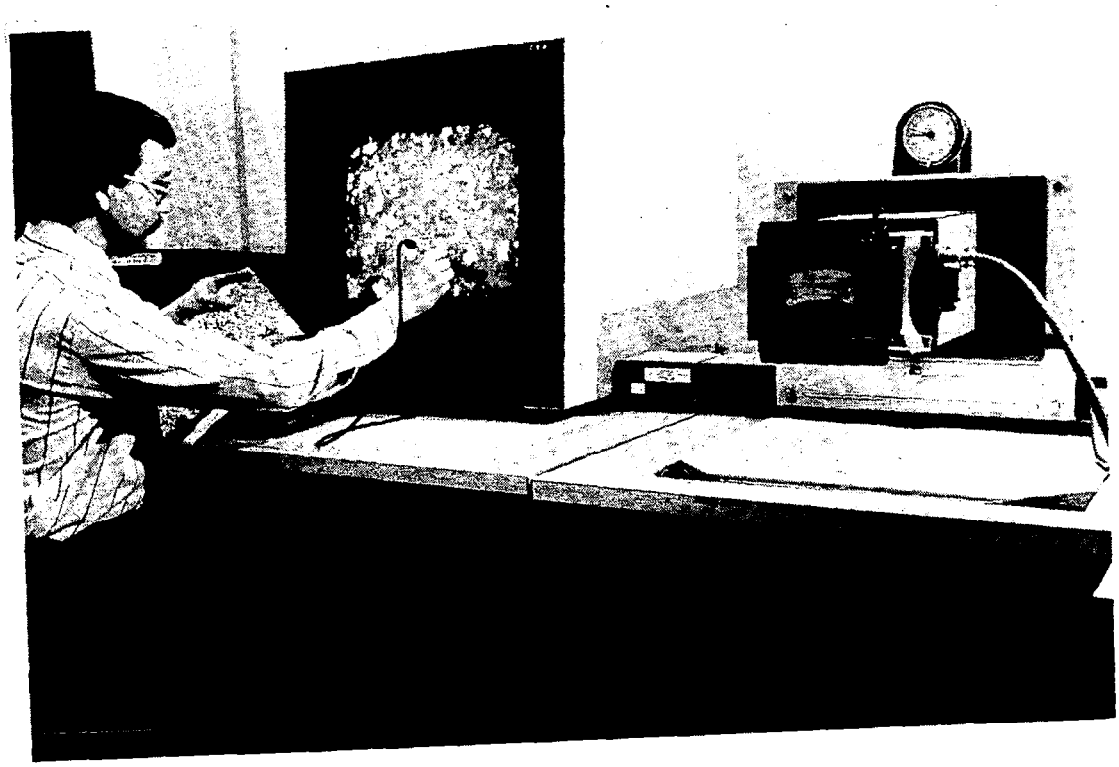


Figure B8. IBM 4507 digital display

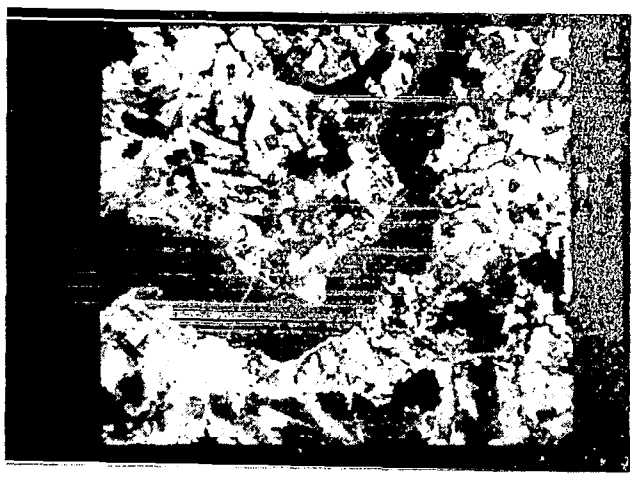


Figure B9. Gray scale image of Landsat spectral band 5, produced on the digital display.

Figure B10. Gray scale image of SAR data, produced on the digital display.

screen capability it may also be used to visually evaluate the registered data set. Also through the use of color filters false color imagery may be produced by the photocopy unit.

Images from two filmwriters were used in this project. The digital microdensitometer at NASA Wallops (see Figure B11) can function as a film scanner or a filmwriter. The microdensitometer is, with minor modifications, constructed from the Optronics Photoscan System P-1000, and the Optronics Photowrite System P-1500. When used as a filmwriter, the Wallops instrument exposes film at 64 repeatable gray levels over an optical density range from background to about 3. The film size when writing is eight inches by ten inches. The aperture may be selected to be 12.5, 25, or 50 micrometers. An image generated on this instrument is shown in Figure B12.

IBM's filmwriter at Gaithersburg, Maryland was also utilized in this project. IBM's filmwriter is not commerically available and is only for use to support contracted projects. An example of imagery written on this device is shown in Figure B13.

The main advantage in using a filmwriter to reconstruct images is the extremely high quality image producible. To some degree the image quality may be attributed to the very high resolution and to the high number of gray levels available. Also, the entire data set may be viewed easily on an image which is permanent. Another advantage of the filmwriter is spatial stability, thus scaled products may be produced with fewer distortions than might be introduced by an imaging device such as the photocopy unit on the LARS digital display.

The major disadvantage is the expense of the filmwriter and its operating costs. Secondarily, because of the very small pixel size, visual ground features location to a pixel accuracy becomes difficult. Although some ground features which were not visible in other imagery may now be visible because of the high image quality.

The filmwriter imagery may be used to assess data quality and in evaluating enchanced data sets. It may also be used to locate ground control points and training and test fields and may serve as a guide to locating features to pixel level accuracies, e.g., in conjunction with a line printer. By superimposing separate images produced by the filmwriter of the multiimage registered set, the registration may be evaluated visually.

Methods

During the process of image reconstruction a mapping of pixel data values to gray levels occurs. Just how this mapping is performed affects the appearance of the imagery produced. An elementary approach is a simple one-to-one mapping; for several reasons this approach is often not desirable. The number of gray levels available on the imaging device may not be sufficient to cover the range of the data. Also if the one-to-one mapping is used the image may have a "washed-out" appearance. This would

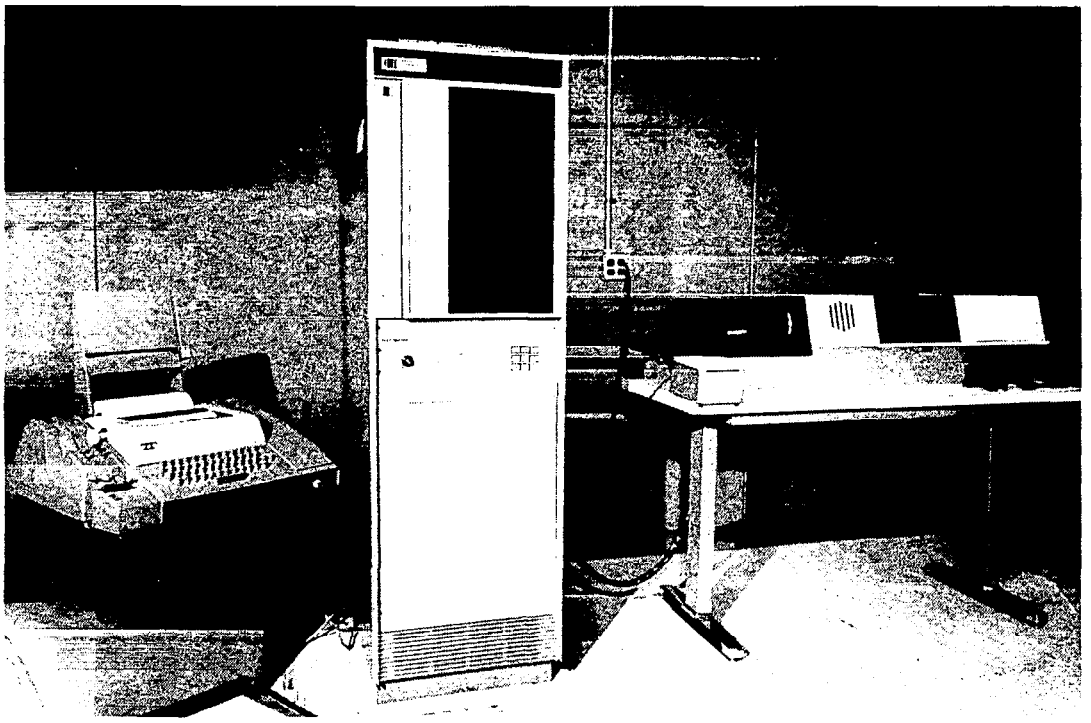


Figure B11. Wallops film scanner and writer

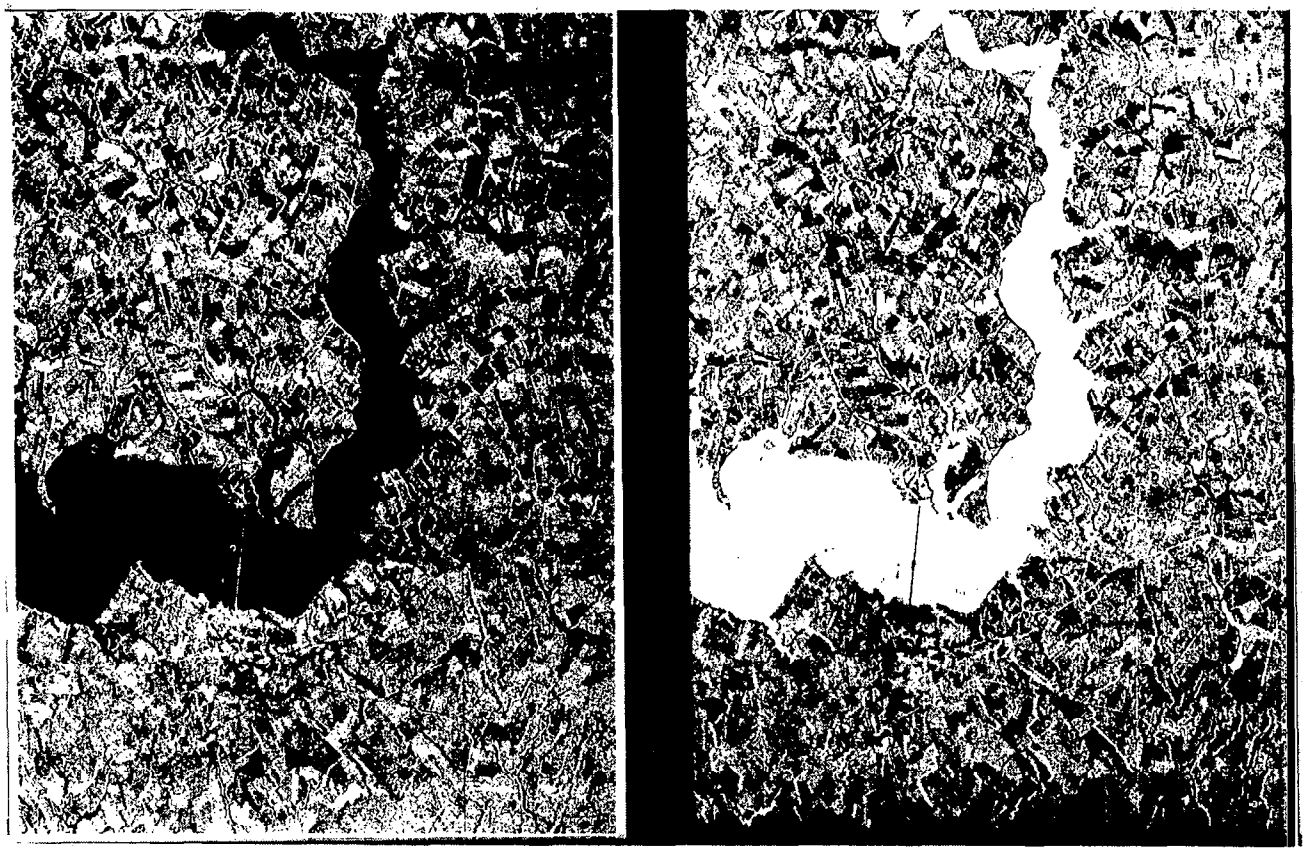


Figure B12. - Gray scale image of SAR data, produced on Wallops film writer

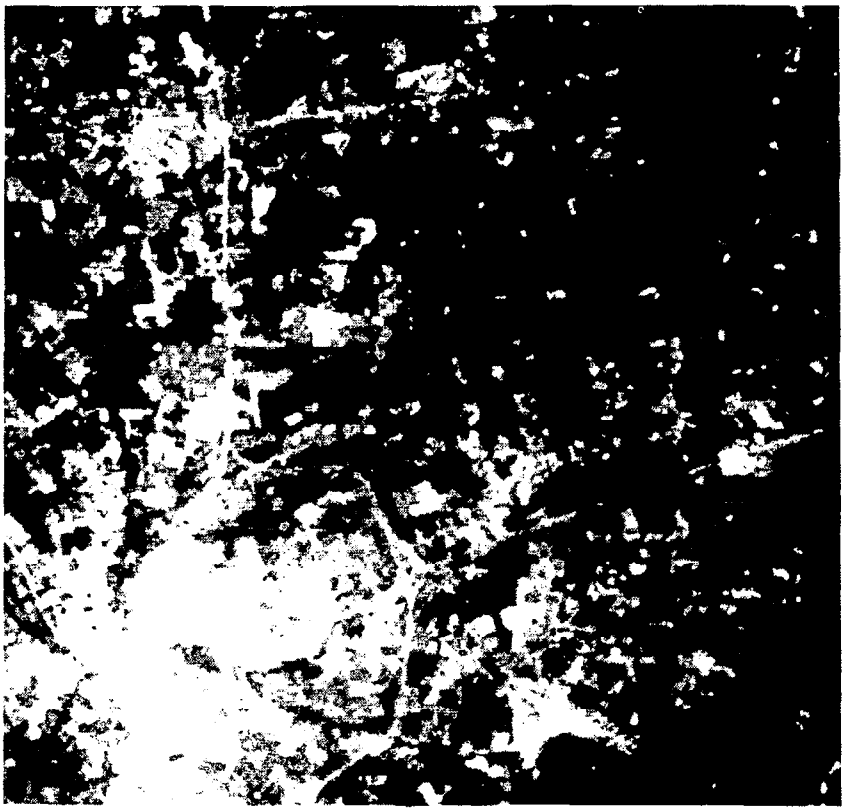


Figure B13. Gray scale image of Landsat spectral band 5 produced on IBM's filmwriter

occur if the data have a Gaussian distribution such that the mid-range gray levels would have a much larger number of data elements assigned to them as compared with extreme gray levels which would have relatively few data elements assigned to them.

One method that may be used to improve the appearance of the image is contrast stretching. This is done when the data values do not fill the gray level range. A small change in the pixel data value is assigned a large change in gray level value. This process may extend over the entire range of the data values or may be used only for a particular set of values. For example, water has fairly low reflectance in the red and infrared channels of Landsat; thus the lowest quarter of the Landsat intensity values might be stretched to fill one half of the gray level range to enhance features in the water. Another method is density slicing where a histogram of the data values is examined, and then ranges of data values are mapped into gray levels such that each gray level will usually have approximately the same number of data elements assigned to them. This method avoids the washed-out appearance and can produce a contrast enhanced image.

Still another method, which may improve the image visually, is to reduce the number of gray levels used. The effects of this method are data dependent since when one reduces the number of gray levels one is also reducing the amount of information available. This may be desirable when the data set is noisy to some degree. For example, when viewing an image with many gray levels, a particular field may appear speckled, while in an image with fewer gray levels, the field may appear as more of a continuous tone, and thus easier to distinguish.

Visual inspection of a histogram of the data set might also be useful in assigning data ranges to gray levels. Here one would look for peaks in the histogram which would correspond to information classes. Ideally, in this manner, an information class wouldn't be mapped into several gray levels and thus blend into neighboring classes, but would be assigned to a single gray level and thus would be distinct. Alternately, a clustering algorithm may be applied to the data set to automatically do this level slicing.

An imaging method that is occasionally of some value is creating a negative image of the data set. The possible benefit of this method may not lend itself to quantitative analysis, yet some features may be more readily recognized in the negative than in the positive image.

The above imaging methods apply to data sets of a single spectral band or to individual bands of multi-band data sets. In imaging a multi-band data set one may wish to view some combination of the bands simultaneously, so that additional information not available in other single spectral bands might be available.

One method of viewing three spectral bands simultaneously is to produce a false color image of the three bands (see Figures B14 and B15). This is done by assigning

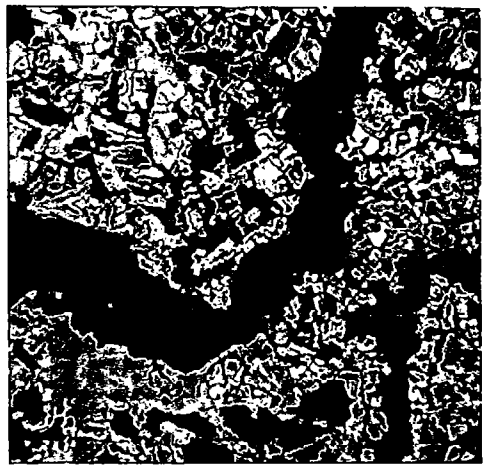


Figure B14. False color image using Landsat spectral bands 4, 5, and 7 produced on the digital display

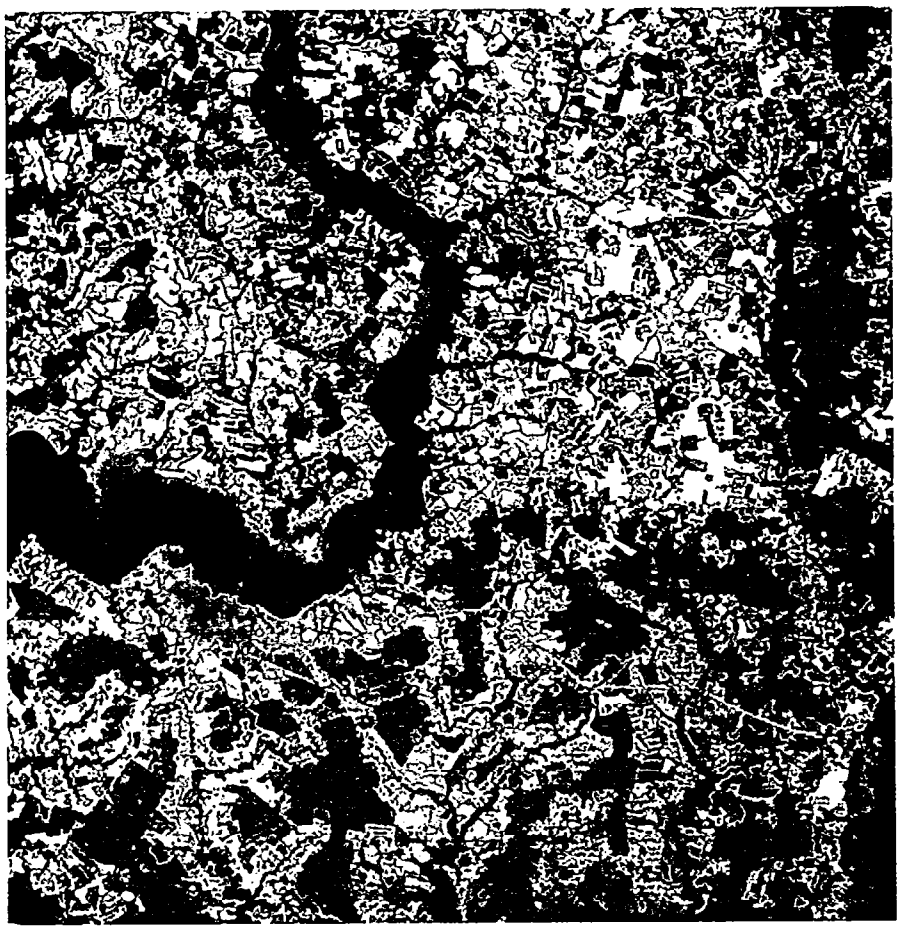


Figure B15. False color image using Landsat spectral bands 4, 5, and 7 produced on IBM's filmwriter

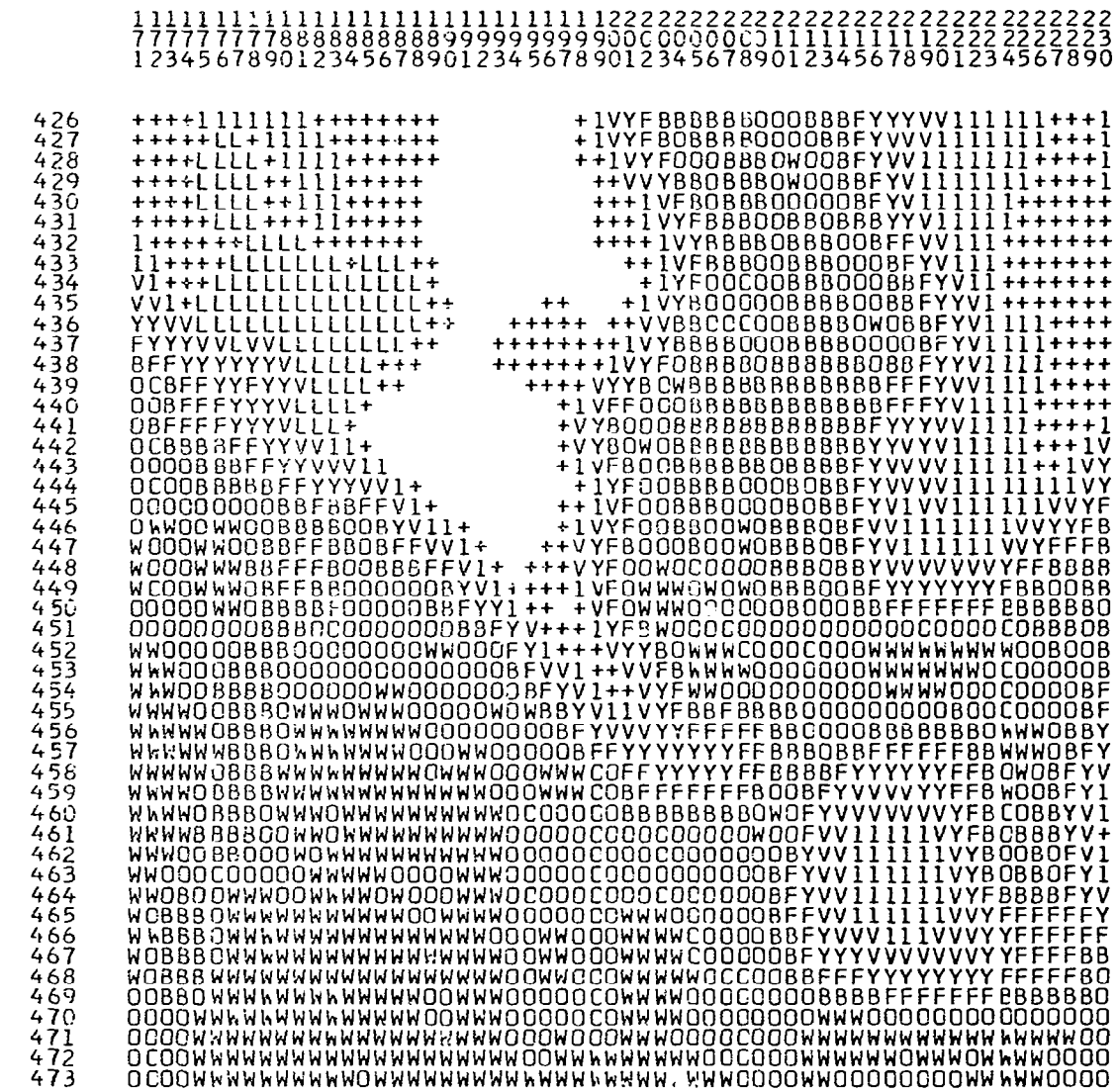


Figure B16. Cluster map of Landsat spectral bands 4, 5, 6, and 7 produced on a line printer

a primary color to each of the three bands to produce the color imagery. One of the aforementioned methods, such as contrast stretching, may be used during the process of creating the false color image. In this manner a feature that does not stand out in a particular band may become more distinguishable.

Another method for viewing several bands simultaneously is to apply a multi-band clustering algorithm to the data set (see Figure B16). Using this method the gray level would be associated with information classes instead of simply representing data response ranges. This in turn could aid in locating ground features to be used as ground control points during the registration process.

Conclusions

During the registration process the type of imagery needed depends on the use that is to be made of the imagery. During registration the primary use of imagery is for locating ground control points. In locating these control points within the data sets there are two qualities the imagery needs to possess. The imagery needs to be produced with small pixels so that the ground features may be recognized, and the resolution needs to be large enough so the features may be located to within a pixel accuracy. To try to accomplish both within a single image is at best a compromise. A more viable solution to this problem is to produce two images, one with small pixels and one with large pixels. Ground control points may then be located in the small pixel image which in turn serves as a guide in locating the control point in the large pixel image with individual pixel accuracy. Producing a good small pixel image requires the use of a filmwriter while a grayscale plotter or a line printer can be used to produce the large pixel image. The grayscale plotter may, depending upon the size and quality of the data set, be more useful than the line printer.

When using the imagery to visually evaluate the registration the capability to simultaneously view multiple bands is necessary. This may be done by superimposing the bands or by producing a "split-screen" image. The superimposed image tends to be more useful since a larger area may be evaluated. To produce the superimposed image one needs to use either a filmwriter or a digital display with a photocopy unit. The filmwriter would have the advantage of producing a less distorted image.

Ideally the imaging devices needed in a registration system would include a filmwriter and a grayscale plotter. The filmwriter would provide the small pixel image needed for recognizing ground features, the capability for producing false color imagery, and the ability to produce superimposed imagery. The grayscale plotter would provide large pixel imagery and economical images that could be used to evaluate the imaging methods and data quality.

APPENDIX C

COMPUTER PROGRAMS FOR SLDMS

This section outlines the computer programs which are necessary to the implementation of the SLDMS. Further information on these programs is available in Reference 10.

SAR Processing Programs

Reformatting Program

This reformatting program converts the Wallops digital microdensitometer magnetic tape format, i.e., the Wallops Image Scanning System (WISS) format to the LARS Multispectral Image Storage Tape (MIST) format. The inputs, functions, and outputs of this computer program are shown in Table C1.

Table C1. REFORMATTING PROGRAM FUNCTIONS

Input	Process	Output
1. Wallops (and other) Image System Formatted Digital Tapes(s)	A. Convert WISS to MIST format	1. MIST format tape
2. User Request	B. Create FORM 17's	2. Printer listings

Shading Correction

This computer program performs along track and across track normalization to reduce shading in SAR imagery. Inputs will be MIST formatted SAR imagery. Outputs will be SAR imagery on tape in the same format. Each correction will be performed independently (Table C2).

Table C2. SHADING CORRECTION PROGRAM

Input	Process	Output
1. MIST format data set tape	A. Perform across track normalization	1. MIST format data set tape
2. User requests	B. Perform along track normalization	2. Printer output

Low Pass Filter Program

Perform low pass filtering on SAR data to reduce speckling in the data (Table C3).

Table C3. LOW PASS FILTER PROGRAM

Input	Process	Output
1. MIST format SAR data set tape	A. Perform low pass filtering of data	1. MIST format SAR data set tape
2. User request		2. Printer output

Distortion Evaluation Program

Perform distortion evaluation based on the acquired corresponding control points form the SAR and Landsat data sets. Lineprinter or GDATA plots should be options. Up to second order functions may currently be evaluated. Appropriate polynomial coefficients are generated. (Table C4).

Table C4. DISTORTION EVALUATION PROGRAM

Input	Process	Output
1. MIST format data set tape(s)	A. Calculate least squares regression polynomial coefficients for parametric and/or affine and/or bi-quadratic polynomial fits	1. MIST format tape(s) 2. Printer output
2. LARS 19 cards		
3. Minimum number of acceptable checkpoints	B. Calculate error estimation for each type of fit	
	C. Create file for lineprinter or printer/plotter output	

Registration/Resampling Program

This program is designed to perform registration of a data set with respect to a reference standard grid (Table C5). In the SLDMS the reference standard grid is expected to be a Landsat data grid which has been corrected to UTM. Linear or biquadratic with linear interpolation registration may be specified. For pixel interpolation, nearest neighbor or cubic interpolation may also be specified.

Table C5. REGISTRATION/RESAMPLING PROGRAM

Input	Process	Output
1. MIST format SAR data set tape	A. Perform registration in linear or biquadratic /linear mode	1. MIST format tape with SAR data
2. User requests data area to be registered; function coefficients to be used; interpolation to be used; reference standard grid	B. Perform pixel interpolation nearest neighbor or cubic	2. Printer listing
	C. Perform 1, 2 on selected subset of data	

Imaging Programs

These programs, which use MIST formatted image data, have been developed for the LARSYS or LARSYSDV systems and are documented as part of those systems. The programs output images on, for example, a Varian plotter, a line printer, and a cathode ray tube (CRT).

Landsat Processing Programs

Reformatting Program

The reformatting program converts Landsat MSS computer compatible tapes (CCTs) into LARS Multispectral Image Storage Tape (MIST) format. The inputs, functions, and outputs of this program are stated in Table C6. It is important to note that the reformatting program performs different functions on each of the different kinds of input data. The different possibilities are shown in Table C7.

Table C6. REFORMATTING PROGRAM FUNCTIONS

Input	Process	Output
1. MSS computer tapes: a. Uncorrected b. Partially processed c. Fully corrected	A. Convert uncorrected MSS data from X format to MIST format.	1. Magnetic tape which contains e.g., image in MIST format and Ancillary Data Set.
2. User's requests.	B. Convert partially processed MSS data from BSQ format to MIST format.	2. Line Printer Listing which contains, e.g., histograms and bookkeeping information.
	C. Convert fully corrected MSS data from BSQ format to MIST format.	
	D. Remove line-length correction (LLC) pixels from uncorrected data.	
	E. Compute histograms for MSS Imagery.	
	F. Compute striping reduction tables.	
	G. Perform striping reduction.	
	H. Extract ancillary data from input tapes.	

Table C7. REFORMATTING PROGRAM OPTIONS

Uncorrected	Partially Processed	Fully Corrected
Convert to MIST	Convert to MIST	Convert to MIST
Remove LLC pixels		
Detector histograms	Detector histograms	
Striping red. tables	Striping red. tables	
Striping reduction	Striping reduction	
Extract ancillary data	Extract ancillary data	Extract ancillary data

Automatic Control-Point Location Program

The automatic control-point location program is used to locate the input-space coordinates of a set of points whose geodetic coordinates are known. The output from this program is a set of pixel coordinates of control points. These coordinates are specified to a fraction of a pixel, and the location accuracy is assumed to be within one tenth of a pixel for a good control point.

The specific functions of this program are shown in Table C8. Basically, the program extracts suitable control points from a library tape, creates corresponding search areas from an image tape, correlates the data, and determines the search-area coordinates of the control points. The tape containing the control points is a Control Point Library Copy Tape, which is a copy of a tape from NASA's control point library. It is also possible to obtain control points from an image tape.

Manual Control-Point Location Program

The manual control point location program has two basic functions:

- To process image data in a way that will enable a user to determine control-point coordinates manually.
- To resample a fully corrected image in order to change its scale (i.e., its pixel spacing).

Each of these functions is described below. The specific functions of this program are listed in Table C9.

In order to assist a user to determine control-point coordinates manually, this program produces two kinds of output. First, it generates a resampled image data set for each control point. Second, it displays a control-point data set on a line printer listing by simulating pixel gray levels with combinations of printer characters including overstrikes.

Table C8. AUTOMATIC CONTROL-POINT LOCATION PROGRAM FUNCTIONS

Input	Process	Output
1. Control Point Library Copy Tape.	A. Extract geodetic control points from:	1. Enlarged ancillary data set containing control point locations.
2. Ancillary data set (from Reformatting).	a. Control Point Library Copy Tape	2. Printer listing of control point locations and other supporting data.
3. MIST format tape (from Reformatting).	b. MIST format image tape	
4. User's requests.	B. Extract search areas from image tape.	
	C. Horizontally resample search areas (and area near control points, if from image tape) to correct high-frequency errors.	
	D. Obtain correlation surface for each control point.	
	E. Fit correlation surface to find control point location to sub-pixel accuracy.	
	F. Evaluate quality of each control point.	

Table C9. MANUAL CONTROL-POINT LOCATION PROGRAM

Input	Process	Output
1. MIST format tape (from Reformatting).	A. Extract small subimages to computer memory.	1. Magnetic tape in MIST format
2. User's requests.	B. Resample small subimages including: a. Scale change b. High-frequency error correction	2. Line printer listings of: shadeprints (picture-prints); checkpoint restart information; histograms; and other supporting information.
	C. Resample large subimage for scale change.	3. Checkpoint restart data set.
	D. Checkpoint restart capability for large subimage resampling.	
	E. Histogram small subimages.	
	F. Obtain translate tables using histograms.	
	G. Print shadeprint (picture-print) on listing using overstrikes.	

The resampling that is performed for a control-point location task has two purposes. First, horizontal geometric distortions that are high-frequency in nature are removed during the resampling process. These distortions are listed in Table C10. Next, a change of scale is performed on the resampled image data set for a control-point. This is done to increase the control-point sub-pixel accuracy level.

Table C10. HIGH-FREQUENCY HORIZONTAL GEOMETRIC ERRORS

Line length variations
Earth rotation errors
Sampling delay errors
Band-to-band offset
Mirror velocity errors

Another basic function of this program is to rescale fully corrected MSS data to the pixel spacing selected by the user. This is a much simpler resampling algorithm than that needed to correct partially-processed or uncorrected data. It can be performed using fewer computations.

Although the two basic functions of this program are quite unrelated, there is a good reason for incorporating them in a single program. Each of these functions involves resampling image data in order to change its scale. The two resampling processes are similar in most respects. Therefore, program development and maintenance will be simplified if these functions are performed by the same program.

Geometric Transformation Program

The geometric transformation program has the following basic functions for uncorrected or partially processed scenes:

- To define the transformation between the corrected output space and the high-frequency-corrected input space.
- To assess the accuracy of the geometric transformation.

The resulting geometric transformation information is added to the ancillary data set for the scene. It is specified as a finite, tabular function. A set of output-space grid points is defined, and the corrected input-space coordinates are determined. This grid-point correspondence is the primary output of the program.

The specific functions of this program are shown in Table C11. For uncorrected data for which no control points exist, models that correct all known systematic errors are used. If control point locations are available for an uncorrected or partially processed data set, models that correct all known systematic and scene-dependent errors are

Table C11. GEOMETRIC TRANSFORMATION PROGRAM

Input	Process	Output
1. Ancillary data set.	A. Systematic error modeling for uncorrected MSS data.	1. Enlarged ancillary data set containing grid-point correspondence.
2. User's requests.	B. Scene error modeling for uncorrected or partially processed MSS data.	2. Printer listing of supporting data.
	C. Output space definition.	
	D. Interpolation grid point creation from error models.	
	E. Interpolation grid point creation from ancillary data (for partially processed scenes).	
	F. Geometric error assessment from error models.	

used. These error models are then used to determine grid point correspondence.

In the case of scene correction, this program also provides an assessment of the geometric errors. This is done using the covariance matrix of the attitude/altitude fitting process, as described in the section LANDSAT MSS IMAGE TO MAP REGISTRATION.

This program can also create a grid point correspondence from the ancillary data for a partially processed scene.

Resampling Program - The resampling program creates a geometrically corrected MSS image data set from either an uncorrected or a partially processed scene. It can resample the data using nearest-neighbor or cubic-convolution resampling algorithms. Horizontal, high-frequency, geometric errors are removed by this resampling program.

The specific inputs, functions, and outputs are shown in Table C12. A checkpoint-restart capability is included because cubic-convolution resampling is a computationally bound process.

Table C12. RESAMPLING PROGRAM

Input	Process	Output
1. Ancillary data set from geometric transformation program.	A. Set up interpolation grid point arrays.	1. MIST format tape.
2. MIST format tape from uncorrected or partially processed scene.	B. Horizontal, high-frequency error correction.	2. Printer listing.
3. User's requests.	C. Cubic convolution resampling.	3. Checkpoint restart data set.
	D. Nearest neighbor resampling.	
	E. Checkpoint restart capability.	

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA RP-1039		

4. Title and Subtitle	5. Report Date
SYNTHETIC APERTURE RADAR/LANDSAT MSS IMAGE REGISTRATION	June 1979
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
H. E. Maurer and J. D. Oberholtzer, NASA Wallops Flight Center P. E. Anuta, Purdue/LARS (Editors)	
	10. Work Unit No.

9. Performing Organization Name and Address	11. Contract or Grant No.
National Aeronautics and Space Administration Wallops Flight Center Wallops Island, VA 23337	
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546	Reference Publication
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The results of a SAR/Landsat MSS data merging system study are discussed. This system study was conducted to determine the algorithms and procedures necessary to merge digital aircraft SAR and Landsat MSS imagery. A plan was then developed from this system study, in which the design of a SAR/Landsat Data Merging System (SLDMS) was pursued. The results from the investigation of three SAR/Landsat MSS data sets indicate that the registration of SAR imagery with Landsat MSS imagery is both feasible, from a technical viewpoint, and useful, from an information-content viewpoint. Multisensor registration is likely to be a desirable process in the future.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Landsat Satellite Multispectral Scanners Pattern Registration Radar Imagery Side Looking Radar	Unclassified - Unlimited STAR Category 43

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
UNCLASSIFIED	UNCLASSIFIED	234	$9.50

* For sale by the National Technical Information Service, Springfield, Virginia 22161

☆U.S. GOVERNMENT PRINTING OFFICE: 1979—638-704